Effects of manner of delivery in on-line pragmatic inferences

Jia E. Loy

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

to

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

2017

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Jia E. Loy

Abstract

In everyday discourse, *the way* a speaker produces an utterance is often as important as the content of the message. Therefore, taking into account their manner of delivery is relevant for successful comprehension. This thesis investigates the influence of a speaker's manner of delivery on a listener's pragmatic inferences, with a focus on the nature and timing of such effects during on-line comprehension.

Previous research has established the role of a speaker's manner in driving listeners' off-line pragmatic comprehension, as well as its effect on their on-line content-based expectations. However, considerably less attention has been paid to its influence on listeners' on-line pragmatic inferences.

The thesis focuses on two general contexts in which pragmatic inferences arise: The comprehension of scalar expressions and the perception of deception. The former has been investigated with a class of scalar implicatures from a time course perspective, but is under-explored with respect to a speaker's manner of delivery; the latter has been studied extensively in relation to manner of delivery, but has yet to be addressed in terms of the timing of the process. We focus on these two areas in attempt to address broader questions about the perceptual relevance of manner of delivery and the time course of pragmatic comprehension.

We approach these questions through a series of comprehension experiments that combine eye- and mouse-tracking technologies. We measured listeners' eye movements and mouse coordinates as they selected between objects on a screen in response to unfolding speech. Critical utterances were either conventional, fluent forms, or characterised by disfluency. Our results demonstrate the immediacy

with which this manipulation influenced listeners' pragmatic inferences, in both the comprehension of scalars and the perception of deception. In the case of deception, this effect is robust even in the face of other sources of information which might modulate the effect. This time course of events has implications for traditional models of language comprehension, which maintain the view that pragmatic comprehension is slow and resource demanding.

In a final experiment, we situated the comprehension paradigm in an ecological, interactive context, to explore the production and perception of manner of delivery in real-time conversation. Here, we found a surprising mismatch between speakers' production and listeners' perception of manner-based cues to deception. Taken together, the results from the studies highlight the role of manner of delivery, as well as broader contextual factors, in shaping the pragmatics of an act of communication.

Lay summary

When understanding the full meaning of a message, listeners often have to attend to a speaker's manner of delivery (*the way* they speak) in addition to the content of the message (*what* they say). This thesis investigates how a speaker's manner of delivery influences a listener's eventual understanding of what is said.

We focus on two examples: listeners' interpretation of *some* (whether the speaker means *some but not all* or *some and possibly all*), and listeners' perception of deception (whether the speaker is lying or telling the truth). We make use of a series of experiments which measure listeners' eye and mouse movements as they select between two objects on a screen whilst hearing a recorded utterance from a speaker. We compare utterances which were fluent to those which included disfluency (e.g., an *um* or *uh*). By analysing where listeners look and move the mouse toward as the utterance unfolds, we can gain an indication of how their interpretation develops depending on the speaker's fluency.

Our results demonstrate that the meaning listeners infer is very quickly influenced by the speaker's manner of delivery. In the case of deception, listeners continue to be affected by this even when other sources of information that they can rely on are available. These results challenge traditional models of language comprehension, which maintain that comprehension going beyond the literal meaning of what is said is slow and effortful for a listener.

In a final experiment, we explore the role of manner of delivery in real-time conversation by adapting the experiment into an interactive game of deception. Here, we find a suprising mismatch between the cues that listeners use to infer that the speaker is being deceptive, and those that speakers actually produce when lying. Taken together, our results show the importance of manner of delivery, as well as broader contextual factors, in shaping an act of communication.

Acknowledgements

I would like to acknowledge several people and funding sources for their support over the course of this PhD.

I am grateful to the University of Edinburgh, College of Arts, Humanities and Social Sciences, for a PhD studentship which helped fund the research in this thesis. A number of research grants from the School of Philosophy, Psychology and Language Sciences, as well as a separate Grindley grant, also supported my attendance at various conferences.

I take this opportunity to thank my supervisors, Martin Corley and Hannah Rohde, for their continued guidance and encouragement, as well as tireless feedback on multiple drafts of work. Above all, I am grateful to them for giving me the confidence to believe in my abilities throughout this PhD.

My fellow PhD students, with whom I've had many insightful conversations, have helped shape my ideas and understanding over the years; in particular I'd like to acknowledge Jo King, Elli Drake, Mirjam Eiswirth and Milan Valášek. I also thank Aine Ito for imparting her wisdom on eye-tracking to me at the start of my PhD. Further thanks to the following people for lending their voices to various experiments: Catriona Gibb, Sam Gibbon, Catherine Crompton, Michela Bonfieni, Lauren Hadley, James Dixon, Ellise Suffill, and Sam Laing. For his endless patience during my early days of learning to program, as well as constant friendship, I am especially grateful to Philip John Gorinski.

Most of all, I thank my family, for always being there despite the distance between us. Without them, I would not have been able to complete this PhD.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Successful comprehension often involves going beyond the surface meaning of what a speaker says. For example, everyday discourse is filled with idioms, metaphors, indirect requests and ambiguous expressions, requiring listeners to reason about what the speaker really means or intends in the given context. This reasoning involves a process of pragmatic inferencing in which listeners draw on information beyond what is said (Grice, 1975). One source that listeners can potentially draw on is the speaker's *manner of delivery*.

This thesis sets out to explore how manner of delivery influences the meaning of an utterance for listeners. We define manner of delivery broadly as a set of properties relating to *the way* an utterance is produced. These may encompass paralinguistic aspects of communication such as a speaker's intonation, rate of speech, accent and fluency, but also more broadly their facial gestures and body language that accompany speech. These features are extrinsic to the lexical-semantic content of an utterance, but may contribute to the meaning of a message in their own right (see e.g., Wagner & Watson, 2010 for a review on prosody).

In this thesis, we focus on manner of delivery in the form of utterance fluency, looking at the difference between fluent and disfluent speech. Disfluencies have traditionally been perceived as a symptom of difficulty encountered by the speaker during planning or articulation (Clark & Clark, 1977). Accordingly, this view regards such interruptions as "noise" that has to be filtered out by the parsing mechanism (Lindsay & O'Connell, 1995). However, recent evidence highlights the perceptual relevance of disfluencies to listeners, who may make use of them to form inferences about the speaker or the discourse (Clark & Fox Tree, 2002). Studies on disfluency comprehension show that filled pauses can guide a listener's on-line parsing of an utterance, creating probabilistic expectations about the upcoming message content during the moment-to-moment processing of the linguistic input (J. E. Arnold, Tanenhaus, Altmann, & Fagnano, 2004).

While the effect of disfluency on a listener's construction of message content is well-established, less is known about its influence on their on-line pragmatic inferential processes. Does this aspect of manner of delivery shape a listener's on-line pragmatic hypotheses about a linguistic message? If so, what is the time course with which these expectations are formed? This thesis aims to address these questions through a series of experiments. The first five experiments employ an eye- and mouse-tracking methodology to investigate listeners' pragmatic inferences during real-time comprehension. By analysing listeners' eye and mouse movements directed toward objects on a screen in response to speech, we can gain an idea of the cognitive processes underlying comprehension of the unfolding linguistic input (Huettig, Rommers, & Meyer, 2011). The sixth experiment situates the lab-based comprehension paradigm in a real-life, ecological context by adapting it into an interactive, dialogue game. This allows us to simultaneously examine the production and perception of disfluencies, as well as other aspects of manner of delivery, while speakers and listeners interact in a controlled, yet relatively natural context.

1.1 Thesis overview

The thesis is structured as follows. In Chapter 2, we review the literature relevant to the scope of the research. We provide a broad overview of manner of delivery from both a production and a perception perspective. This chapter highlights that speakers may encode pragmatic information through their manner of delivery, and moreover, that listeners' pragmatic inferences are driven by a speaker's manner-based cues. The question of when this information is integrated is situated within two competing models of language comprehension—the *standard pragmatic model* and the *one-step model*. This is followed by an overview of eye-tracking and mouse-tracking, with a focus on how these methods are incorporated within the visual world paradigm. This methodology is employed by the experiments in the thesis.

Chapters 3–6 detail the empirical work of the thesis, which explores how a speaker's manner of delivery influences a listener's pragmatic inferences. We focus on pragmatic inferencing in two contexts: The comprehension of scalar expressions and the perception of deception. In Chapter 3, we present a study investigating the effect of manner on the interpretation of the scalar quantifier *some* (Experiment 3.1). Consistent with previous work, we find that listeners tend to derive the pragmatic inference of *some* to mean *not all*; however, a disfluency in the utterance attenuates this bias in favour of a literal (*possibly all*) interpretation. Importantly, this difference unfolds rapidly in listeners' eye and mouse movements, demonstrating that pragmatic inferences occur on-line, and are modulated immediately by a cue in the speaker's manner.

In Chapter 4, we turn our attention to the perception of deception. We present a study investigating the effect of a speaker's manner on listeners' pragmatic inferences about a speaker's truthfulness. Our findings show that disfluency biases listeners toward perceiving the speaker to be lying. As with Experiment 3.1, this disfluency bias emerges from the earliest moments of comprehension, and furthermore holds regardless of whether the speaker is disfluent at the start of the utterance (Experiment 4.1) or in the middle (Experiment 4.2). The latter finding has implications for disfluency theories which propose separate meanings for utterance-initial and utterance-medial disfluencies (e.g. Clark & Fox Tree, 2002).

Chapter 5 extends findings from Chapter 4 by investigating the influence of speaker-specific expectations on listeners' biases to associate disfluency with lie behaviour (disfluency-lying bias). We examine two forms of speaker expectations: Expectations based on a speaker's perceived veracity (Experiment 5.1) and expectations based on a speaker's accent (Experiment 5.2). In both cases, we find that the disfluency-lying bias is robust to the influence of listeners' inferred expectations relating to a speaker. This may suggest that disfluency is a "universal" cue to deception, reflecting the ingrained nature of such stereotypes of deceit (cf. Zuckerman, Koestner, & Driver, 1981).

In Chapter 6, the final experiment, we explore the disfluency-lying bias within a real-life, ecological context.[1] We present a study that adapts the visual world comprehension paradigm from our earlier experiments into an interactive dialogue game of deception. This paradigm allows us to examine the real-time production and perception of various manner-based cues in a controlled, yet naturalistic setting. Here, we find an unexpected result in the form of a mismatch between speakers' production of and listeners' expectations surrounding manner-based cues that index lie behaviour.

Finally, in Chapter 7 we discuss the findings from the empirical chapters and present the conclusions of the thesis.

[1]For expository purposes, we present the experiments in this order in the thesis; note, however, that they were run in the following order: 4.1, 4.2, 6.1, 5.1, 5.2, 3.1.

Chapter 2

Background

Language comprehension is a complex act of cognition. The apparent ease with which humans understand language belies a sophisticated set of mechanisms which map meanings onto utterances. The goal of this process is not merely to grasp the semantic content of sentences, but to interpret what speakers may mean or intend within the broader discourse context (Grice, 1975). Moreover, natural, everyday speech unfolds rapidly (3–5 words per second; Picheny, Durlach, & Braida, 1986), and is often noisy and prone to error (Shriberg, 2001). Successful understanding thus requires the ability to navigate complex input and decode meaning on a short time scale.

How, then, do comprehenders make use of the rapidly-unfolding input to construct a meaningful, coherent message? In this thesis, we explore how listeners draw on *the way* something is said to map meaning onto *what* is said. In particular, we are interested in the construction of meaning beyond the information explicit in the utterance. Grice (1975, 1989) proposed that such meanings are derived through a process of pragmatic inferencing by which listeners reason about what a speaker may intend over and above the surface form of what they say (cf. Searle, 1979). In the next section, we explain what this process entails for listeners, and how it affects the meaning they derive.

2.1 Literal meaning vs. pragmatic meaning

In linguistic theories of how meaning is encoded, a distinction is often made between the conventional meaning of the individual words, and the contributions made by the wider context surrounding the discourse. This distinction has roots in the Gricean notion of conversational cooperation (Grice, 1975), which assumes that speakers strive to be rational and to make purposeful contributions to the discourse (the *Cooperative Principle*). Grice's framework is achieved via tacit adherence to the four maxims that govern communication:

Quality	Do not say what you believe to be false
Quantity	Make your contribution as informative as is required
Relation	Be relevant
Manner	Avoid obscurity of expression

(Grice, 1975, p.45–46)

Grice further notes that while speakers are generally bound to the Cooperative Principle, the maxims can be exploited to communicate more than what is actually said. For example, a speaker may express irony through an utterance such as "What lovely weather we're having" on a stormy day (a clear violation of the first maxim of Quality). In order to maintain the assumption of cooperation here, the listener is led to infer that the speaker does not actually mean what they have literally said, but may have used the expression to convey a particular attitude, such as contempt or ridicule toward the weather (Clark & Gerrig, 1984).

The fact that listeners reason about what a speaker really means (but indirectly communicates) has given rise to a principled division between two forms of interpretation—literal and pragmatic—from which distinct meanings are said to arise. Grice (1975) distinguishes between these two levels of meaning as *what is said* and *what is meant.* The literal meaning, or *what is said*, refers to the "core" meaning embedded in the surface word forms, and how they combine into a larger whole. The literal interpretation of "What lovely weather we're having", for example, is a straightforward expression of a state of good weather. This meaning is context-independent, and holds regardless of the actual state of the weather.

The pragmatic meaning is concerned with what a speaker may mean or intend beyond what they actually say. Pragmatic interpretation involves an inferential process via which listeners draw on information from the wider context to derive an appropriate meaning (Grice, 1975, 1989; Searle, 1979). This information may be linguistic, such as a speaker's intonation or choice of expression (e.g., Brennan & Williams, 1995; Glucksberg, 2001), or non-linguistic, such as the listener's knowledge about the speaker or the world (e.g., Hagoort, Hald, Bastiaansen, & Petersson, 2004). In fact, almost any type of information can in principle be relevant to pragmatic understanding. In the case of "lovely weather" on a stormy day, for instance, the listener may test the utterance against the situational context (the immediate environment) to deduce a pragmatic interpretation of ironic intent. By the same token, the type and nature of inference that the listener can derive is in principle unbounded, constrained in reality only perhaps by the contextual information available. One could, for instance, conceive an alternative scenario in which a listener infers, based on prior knowledge that the speaker enjoys walking in the rain, that the speaker is expressing genuine joy about the weather. In sum, while the literal meaning of an utterance refers specifically to the fixed meaning of its components, its pragmatic meaning is definitionally broad, encompassing a range of interpretations that may arise based on the any aspect of the wider context. In this thesis, we focus on two types of pragmatic interpretation: Scalar inferences, where meaning is ambiguous, and the perception of deception, where meaning may be the opposite of what is said.

2.2 Manner of delivery and the construction of meaning

In everyday discourse, many factors can influence the meaning derived by a listener. In particular, in the case of pragmatic interpretation, *the way* a message is communicated often bears weight on *what* is said. These aspects of communication, such as a speaker's intonation, rate of speech, facial expression or gestures, can be collectively viewed as a speaker's *manner of delivery*. These properties all have the potential to contribute information beyond the surface meaning of the words, altering the overall meaning of the utterance. The sentence "What a great idea!", for example, would be interpreted differently depending on whether the speaker produces it in an excited or a sarcastic tone of voice. Taking into account the manner of delivery is therefore relevant to deriving a contextually appropriate pragmatic interpretation.

One way the manner of delivery of an utterance can vary is in its fluency. Although one often imagines communication in the form of fluent, well-formed sentences as typically portrayed by movies or novels, spontaneous spoken discourse is rarely smooth and unbroken, but rather littered with pauses, repetitions, false starts and the like. These disruptions to speech, known collectively as disfluencies, are defined by Fox Tree (1995, p.709) as "phenomena that interrupt the flow of speech and do not add propositional content to an utterance" (see also Shriberg, 1996 and Hartsuiker & Notebaert, 2010 for studies on the various types of disfluencies). Studies of spontaneous speech have shown that approximately 6 out of every 100 words produced by speakers are affected by some form of disfluency (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). Of these, the most common form for many is perhaps the filled pause (*um, uh*; cf. O'Connell & Kowal, 2004). Fox Tree's definition also highlights a key characteristic of disfluencies: that they do not alter the literal meaning of an utterance. However, as we will demonstrate, in changing *how* the content of the utterance is conveyed, the pragmatic interpretation of the message may change.

This section provides an overview of manner of delivery and its effects on the process of meaning construction. We first look at how manner of delivery varies in production, followed by a discussion on its perceptual relevance in comprehension. We focus on manner of delivery in the form of utterance fluency, as this is the aspect investigated by the studies in this thesis. We note, however, that this is just one of many ways in which a speaker's manner may vary, each of which may play a distinct role in shaping the meaning of the utterance.

2.2.1 How does manner of delivery vary in a speaker's production?

Many factors can affect a speaker's manner of delivery. A speaker may adjust their pitch, speech rate, or even facial gestures to converge towards that of their interlocutor (Giles, Mulac, Bradac, & Johnson, 1987; Louwerse, Dale, Bard, & Jeuniaux, 2012; Staum Casasanto, Jasmin, & Casasanto, 2010). They may emphasise their regional accent or dialect to maintain social distance from an out-group member (Babel, 2010; Bourhis & Giles, 1977). Or they may produce a disfluency because they run into difficulties formulating what they want to say (Beattie & Butterworth, 1979; Bortfeld et al., 2001).

Cognitive load is frequently implicated as a factor in a speaker's production of disfluency. One of the earliest accounts by Goldman-Eisler (1968) highlights a relationship between cognitive effort and silent pause production. Goldman-Eisler asked participants to describe and to interpret a series of cartoons, reasoning that interpreting the meaning behind cartoons would be more effortful, leading to an increase in speakers' disfluency production. As predicted, participants paused more frequently and for longer durations when interpreting than when describing, outlining a causal connection between cognitive effort and speaker disfluency.

Although Goldman-Eisler's findings were restricted to silent pauses, her work has served as a base for subsequent researchers to explore the link between speech difficulty and disfluency production. These studies provide consistent evidence that disfluencies index an increased mental load. Studies on the distributional properties of disfluencies show that they tend to occur in more complex turns (Bortfeld et al., 2001), before longer utterances (Oviatt, 1995), near the beginning of sentences (Boomer, 1965; Shriberg, 1996), and at major discourse boundaries, where presumably greater demand is placed on the speech planning mechanisms (Swerts, 1998; Watanabe, 2002). Speech task complexity is frequently cited as an important predictor of a speaker's production of filled pauses (e.g., Barr, 2001; Christenfeld, 1994; Merlo & Mansur, 2004; Siegman & Pope, 1966; Womack et al., 2012). Christenfeld (1994) manipulated the effort demanded of participants describing a path through a maze by varying the number of possible routes at predetermined points in each maze. He found that mazes with more alternate routes elicited a higher frequency of *ums*. These results echo Siegman and Pope's (1966) findings that speakers produced more filled and silent pauses when describing Thematic Apperception Test (TAT) cards that were highly ambiguous compared to cards that were less ambiguous. Studies such as these demonstrate a relationship between cognitive load and disfluency, with greater load associated with increased disfluencies.

Cognitive load alone, however, does not appear to fully explain a speaker's disfluencies. Bortfeld et al. (2001) observed that while speakers produced more repetitions and restarts with during a harder task (describing abstract geometric figures vs. pictures of children), their production of filled pauses tended in the opposite direction. Moreover, they encountered cases of standalone filled pauses unaccompanied by any speech, such as in the following example:

(1) D: but it would be scrunched down more and the second diamond is
out in front like for feet... does that make any sense... the head...
M: *uh*
D: it's a smaller figure

In a similar vein, Smith and Clark (1993) observed in a task eliciting spontaneous answers to factual questions instances of filled pauses that preceded single-word answers, for example:

(2) Q: In which sport is the Stanley cup awarded?
A: (1.4) *um* (1.0) hockey

Such cases undermine the explanation that filled pauses arise solely as a result of problems in speech planning or articulation. Rather, Bortfeld et al. (2001) suggest that in contrast to other forms of disfluency, speakers may use filled pauses for interpersonal coordination purposes during conversation to facilitate communication. From a conversation analytic perspective, for instance, pauses such as the one in (2) have been attributed to providing a speaker time for planning while simultaneously cueing the listener to tune in (Beňuš, 2009; Clark & Fox Tree, 2002). The notion that such verbalisations may serve as conversation management devices raises questions of how speakers may exploit this function, as well as what pragmatic meanings can be conveyed via disfluency.

Are disfluencies deliberately produced?

The idea that some types of disfluency may perform certain communicative functions stems from the analysis of discourse from a listener-oriented perspective. Research suggests that such cues may be uttered with the intended purpose of signalling pragmatic information to one's addressee (e.g., Clark, 1996, 2002; de Leeuw, 2007; Fox Tree & Clark, 1997; Kjellmer, 2003). An early account by Maclay and Osgood (1959) proposed that some disfluencies may be produced as part of a speaker's attempt to keep the floor, by way of telling the listener not to interrupt. They described this communicative strategy as "some kind of signal ([m, er], or perhaps a repetition of the immediately preceding unit) which says, in effect, 'I'm still in control—don't interrupt me'" (1959, p. 41). Clark's (1996; 2002) account of dialogue similarly proposes that speakers design certain hesitation disfluencies, including filled pauses, repetitions and prolongations, to achieve communication goals. Clark's account rests on the basis that conversation is a joint activity, involving cooperation and coordination between interlocutors. A part of this coordination invokes the use of metacommunicative signals through the speaker's manner of delivery to convey discourse-relevant messages to one's interlocutor. For instance, a speaker who encounters a temporary delay due to difficulty in formulating what to say may produce a filled pause in attempt to keep the floor, as a way of telling the listener not to interrupt, for example:

(3) th- there is a (0.2) a *uh* (0.5) a potential problem,

Clark (2002) noted that the speaker in (3) also cliticised the filled pause onto the preceding "a" (rhyming with "day") such that the two syllables were realised as a single prosodic word. This was a feature that occurred frequently with introductory conjunctions such as "and", "but" and "so" in the corpus. Clark hypothesised that such cliticised forms are used by speakers to signal delay by making a preliminary commitment to the upcoming constituent (in this case in initiating an indefinite noun phrase) but simultaneously conveying an immediate suspension of speech while they formulate what to say next.

Consistent with this view, Kjellmer (2003) observed in a corpus-based analysis on the localisation of filled pauses that *er* and *erm* both frequently co-occurred with conjunctions, and most frequently before the word *and*. Kjellmer proposed that filled pauses function to introduce what he termed *thought units*—"elements that require some deliberation, some planning, which may range from very simple, such as finding an appropriate word, to quite complicated, such as deciding on which out of a great number of facts to communicate, and in what order" (2003, p.174). These information units require preparation, and a speaker who encounters temporary difficulty but is unwilling to yield their turn may rely on a filled pause as a turnholding device in order to keep the floor and avoid being interrupted.

One potential reason for speaker difficulty relates to speaker knowledge. Studies indicate that speakers make use of manner-based cues to convey pragmatic information about their knowledge state. Smith and Clark (1993) investigated the relationship between a speaker's manner of delivery and their confidence about what they were saying, or their *feeling-of-knowing* (FOK; J. T. Hart, 1965). In their study, participants were tape-recorded during a verbal interview in which they tried to recall the answers to general knowledge questions (e.g., "What is the capital of Chile?"). They were subsequently asked to rate on a seven-point scale how confident they were that they would be able to recognise the correct answer (their FOK), and finally tasked with choosing the correct answer to each question in a multiple choice test.

Smith and Clark examined the manner in which participants expressed their responses in the verbal interview, and found that this varied with their FOK judgement for each question. Answers produced with filled pauses, hedges such as "I guess", and rising rather than falling intonation were associated with a lower FOK reported by participants, suggesting that speakers made use of these devices to indicate uncertainty in their answers. An analysis of filled pauses also showed that these correlated with speaker delay: The average latency to respond was 5.06s longer for answers accompanied by filled pauses than those without, and longer delays were associated with *ums* compared to *uhs*. Smith and Clark concluded that participants made use of filled pauses to account for their delay in responding by signalling to the questioner that they were still actively engaged in thought, with *uh* used to signal a minor retrieval problem and *um* a major one.

Separate functions for *um* and *uh* based on depth of retrieval difficulty has also been asserted by Clark and Fox Tree (2002), who similarly proposed that *uh* is used when speakers expect a minor problem, and *um* a major one. Their claim was supported by a comparison of the delays following each variant across four corpora of spontaneous speech, which revealed that *um* was followed by silent pauses more often than *uh*, and the duration of these pauses were significantly longer after *um* than *uh*. Their results echo those of Smith and Clark (1993), as well as Barr (2001), who observed in a referential description task that longer hesitations followed *um* than *uh*. Distinction is also implied with respect to disfluency location: Clark and Fox Tree found that *um* occurred most frequently at the start of intonation units, which they attributed to global, speech-planning difficulties (cf. Brennan & Williams, 1995). This stands in contrast to phrase-medial filled pauses, which are attributed to localised, lexical access difficulties such as word-searching (Beattie & Butterworth, 1979). Clark and Fox Tree concluded that speakers monitor their speech plans for imminent delays, and upon anticipating one they signal this through an *um* or *uh*, the choice of which depending on the nature of the expected delay.

Clark and Fox Tree's assertion has, however, been challenged by O'Connell and Kowal (2005), who found that not only were the majority of filled pauses in their corpus not followed by any delay, those that were showed no difference in duration between the two variants. Moreover, Clark and Fox Tree's basis for their claim in reasoning that "a priori, *uh* and *um* must have distinct causes, just as any two options in behavior do" (2002, p.75–6) is problematic, in that the existence of two distinct options does not necessarily implicate two distinct causes from which the options arise (an example being non-systematic free variation, where speakers alternate between linguistic forms "in unpredictable patterns not constrained by any independent factors"; Davydova, 2011, p.105). Therefore, while evidence suggests that the distribution of *um* and *uh* may vary systematically in speech, it is difficult to establish whether this variation is due to intentional design on a speaker's part, or indeed, whether listeners are sensitive to any (potential) distinction in the signalling function of the two variants. This issue is addressed in the experiments presented in Chapter 4, which afford a comparison of listeners' perceptions of an utterance-initial *um* and an utterance-medial *uh* in the same context of perceiving deception.

It should additionally be noted that it is unclear the extent to which volition plays a role in a speaker's production of manner-based cues. Underlying the hypothesis that speakers design their disfluencies as listener-oriented signals is the understanding that these verbalisations are subject to a speaker's control. Indeed, evidence from production studies indicates that speakers possess a good deal of control over their disfluency production: When offered a monetary incentive to eliminate disfluencies from their speech, Siegel, Lenske, and Broen (1969) observed that speakers were able to reduce their disfluency production to near-zero levels (cf. Boomer & Dittmann, 1964). This finding provides surface evidence that, at least in some situations, speakers are able to deliberately suppress their production of disfluencies. However, it does not necessarily follow that control over one's disfluencies results in conscious design of these disfluencies. Moreover, the ability to suppress an *um* or *uh*, as these studies demonstrate, likely invokes a different process to actually producing them in response to communicative demands. The notion that disfluencies are produced as listener-oriented signals is undermined by findings that these verbalisations are still present in monologue situations (Lickley, 1995), in both naturalistic settings (Schachter, Christenfeld, Ravina, & Bilous, 1991; Tsuchiya, Kogure, Nishizaki, Ohta, & Nakagawa, 2008), as well as experimental task conditions (Broen & Siegel, 1972; Finlayson & Corley, 2012; Swerts, 1998). Finlayson and Corley observed, for instance, that participants were no more disfluent when naming pictures in a dialogue condition, where a confederate interlocutor was present, than in a monologue condition, where no interlocutor was present. Such results challenge a communicative account of disfluencies, which proposes that these verbalisations are produced with intent by speakers to manage listener expectations. Perhaps by way of compromise, Kjellmer suggests that cues such as filled pauses operate "partly under the level of consciousness", and while they may serve to facilitate spoken interaction, "the speaker may well make use of them unconsciously in aiming for a certain effect that s/he achieves without understanding how it was brought about" (2003, p.181).

Ultimately, regardless of whether speakers purposefully encode pragmatic information in their manner of delivery, from a communication perspective a more relevant question is whether listeners make use of the signal potential of these manner-based cues. In the next section we turn our attention to the role of manner of delivery on a listener's perception, and how this affects the meaning they derive from the utterance.

2.2.2 How does manner of delivery influence a listener's interpretation?

Language comprehension studies show that in interpreting an utterance, listeners take into account the speaker's manner of delivery. These studies demonstrate that listeners are sensitive to a speaker's production of manner-based cues, and moreover, that these cues may be used to shape their inferences about the speaker or the discourse. This section provides an overview on manner of delivery as a signal to listeners, looking at a) how it affects a listeners' off-line comprehension of an utterance, followed by b) how it influences their on-line hypotheses as the utterance unfolds.

How does manner influence off-line interpretation?

Fundamental to the perceptual relevance of a speaker's manner of delivery is that listeners are sensitive to this information when interpreting an utterance. After all, if listeners did not attend to the way an utterance is produced, the cue potential of any aspects of its manner would be lost on them. Hence, it is important to first establish that listeners are sensitive to changes in a speaker's manner of delivery.

It is well known that suprasegmental features of an utterance, such as a speaker's pitch or loudness, can serve various signalling purposes for a listener (Lehiste, 1996). For example, the sentence "It looks like an X" can evoke different pragmatic meanings depending on the prosodic contour of the utterance (Kurumada, Brown, & Tanenhaus, 2012).

(4) a. It looks like a ZE bra
H* H* L- L%
b. It LOOKS like a ze bra
L+H* L- H%

A canonical accent placement on the sentence, such as in (4a), elicits the interpretation that the speaker is describing the target noun (in this case a zebra). However, when the verb "looks" was lengthened and emphasised with a contrastive accent as in (4b), Kurumada et al. found that it triggered a bias toward inferring that the speaker was referring to something else (e.g., an okapi, a four-legged animal with black and white stripes only on its legs). More generally, studies have demonstrated that listeners extract information based on the pitch or accent of an utterance, and adjust their interpretation accordingly (Dennison & Schafer, 2010; Ito & Speer, 2008; Nooteboom & Kruyt, 1987; see Cutler, Dahan, & Van Donselaar, 1997 for a review). However, whether or not they are similarly sensitive to a speaker's disfluencies has been debated.

The notion that listeners may make use of a speaker's disfluencies detracts from the traditional view of language comprehension, which has largely treated these verbalisations as noise that is filtered out of the speech stream (Lindsay & O'Connell, 1995). This view stems from early production accounts, which have often associated disfluencies with defective speech, arising as a result of performance-related errors on the speaker's part (Chomsky, 1965). Under this view, disfluencies, such as filled pauses, are a symptomatic by-product of problems encountered by the speaker during the planning or articulation process (the *symptom hypothesis*; cf. de Leeuw, 2007). Consequently, the tacit assumption is that these disfluencies present obstacles to successful comprehension, requiring the parser to edit them out of the signal in order to make sense of the utterance (cf. Levelt, 1989).

Support for the view that listeners filter disfluencies out of the speech signal comes from studies which demonstrate that they frequently go unnoticed during comprehension. Lickley (1995) had participants listen to monologue recordings of speech that included several forms of disfluencies, and compare each recording to a transcript of the speech. The transcripts were edited to remove all disfluencies, and participants were asked to mark out points at which the speech differed from the script. Lickley found a poor detection rate for all types of disfluencies, ranging from 65.4% for filled pauses to 27% for repetitions. In a similar vein, Lickley and Bard (1996) found that listeners asked to transcribe utterances frequently missed disfluencies or failed to transcribe disfluent utterances accurately, with errors clustering around points at which the interruption occurred. These results make a case for the disruptive nature of disfluencies, lending weight to the argument that they are excised by the parsing mechanism to make sense of what is being said.

More recent evidence, however, challenges this view. Research has demonstrated that not only are listeners sensitive to the presence of disfluencies in an utterance, in some cases these verbalisations may facilitate the process of comprehension. Building on Clark's (1996) account of dialogue, Brennan and Schober (2001) hypothesised that some types of disfluencies may present information that aids listeners in understanding speech. They manipulated the fluency of instructions that participants followed to select geometric objects on a graphical display. Disfluent instructions included mid-word interruptions ("Move to the yel- purple square"), some of which were marked by a filled pause ("Move to the yel- uh, purple square") or replaced by a silent pause of the same length. They found that in both cases, listeners were quicker to select the target object (the purple square) compared to when the instructions were fluent. Brennan and Schober proposed that listeners continuously monitor the speech input for cues that the speaker is having problems (cf. Levelt, 1989), using these cues to make repairs incrementally when they detect an error. Their findings suggest that in some situations, such as when dealing with erroneous speech, filled and silent pauses can facilitate the comprehension process, and help listeners to identify upcoming content. A similar result was observed by Fox Tree (2001), who showed that listeners were quicker to recognise a target word in a sentence when preceded by an *uh*, although interestingly no comparable effect was found for words preceded by an *um*. Fox Tree cited this as evidence that *uh* and *um* may have distinct functions in signalling information to a listener (cf. Clark & Fox Tree, 2002).

The above studies provide evidence that disfluencies, in particular filled pauses, may facilitate a listener's comprehension. However, evidence for the communicative potential of these cues remains weak. Although the experiments show that listeners can strategically benefit from disfluencies, this effect may be simply due to a perceptual mechanism, such as by heightening the listener's attention to the upcoming speech (cf. Fox Tree, 2001). It remains unclear whether such cues serve a pragmatic function by signalling information that goes beyond the surface word meaning to listeners.

Brennan and Williams (1995) address this question directly by investigating whether listeners' judgements of a speaker's confidence would vary based on the speaker's manner of delivery. In their study, participants heard recorded utterances of speakers' answers to general knowledge questions (cf. Smith & Clark, 1993), and responded by rating how likely they thought the speaker knew the correct answer to the question, or their Feeling-of-another's-knowing (FOAK). The recordings were digitally edited to create disfluent versions that contained either filled or silent pauses. Based on Smith and Clark's (1993) findings that speakers signal uncertainty through their disfluencies, Brennan and Williams (1995) hypothesised that listeners would draw on this information to form inferences about a speaker's state of knowledge. As predicted, participants gave lower FOAK ratings to disfluent utterances compared to their fluent counterparts, with the lowest ratings observed for utterances containing a filled pause. A post-test interview further showed that a large proportion of participants (76.3%) indicated that they attended to speakers' pauses, outlining an explicit link between a speaker's manner of delivery and a listener's metacognitive assessment of the speaker.

Brennan and Williams's findings were subsequently extended by Swerts and Krahmer (2005), who showed that listeners' FOAK judgements were sensitive to a variety of manner-based cues in both the auditory and visual modalities. Listeners also gave more accurate FOAK ratings (based on FOK ratings that speakers provided in an earlier production experiment) when they had access to both auditory and visual cues, compared to when they only had access to cues in one modality. This result illustrates how multiple features of a speaker's manner of delivery may simultaneously contribute to shaping a listener's interpretation—an aspect of communication often overlooked by laboratory-based paradigms. Although the experiments in this thesis focus on the perceptual relevance of speaker's disfluencies, the study presented in Chapter 6 attempts to address this issue by exploring the production and perception of a range of manner-based cues in both the auditory and visual modalities.

How does manner influence on-line comprehension?

The studies outlined above demonstrate that listeners are sensitive to cues in a speaker's manner. Moreover, these aspects of delivery can not only facilitate content-based comprehension, but also drive pragmatic inferences beyond the surface meaning of the words. However, these studies are limited in how they can inform us about the process by which manner of delivery influences listener comprehension. For example, one explanation for the finding that listeners are faster to identify targets following disfluent utterances (Brennan & Schober, 2001; Fox Tree, 2001) could be due to the temporal delay afforded by the disfluency, providing listeners with more time to select the appropriate response (Corley & Hartsuiker, 2011). Alternatively, a disfluency may signal information to the listener in a more specific fashion, such as a message to the effect that the speaker expects a disruption (cf. Clark & Fox Tree, 2002), allowing listeners to adjust their expectations ahead of time. If the latter explanation were the case, a related topic of interest would be the temporal process with which this information becomes available.

More recent studies have attempted to investigate the real-time process by which disfluency influences comprehension by using paradigms that measure on-line processing. Using an eye-tracking comprehension paradigm, J. E. Arnold et al. (2004) examined how the fluency of an article ("thee uh" vs. "the") influenced a listener's interpretation of the noun that followed. In their study, listeners manipulated objects on a visual display in response to recordings of spoken instructions (e.g., "Put the grapes above the candle. Now put *the/thee uh*..."). On critical trials, displays included two objects with an overlapping phonological onset (e.g., candle-camel), one of which was *given* as the speaker had referenced it in the first part of the instructions (candle), while the other was *new* in that it had not been previously mentioned (camel). Based on evidence from production studies that speakers are more likely to be disfluent when referring to something new than something given (J. E. Arnold & Tanenhaus, 2011), J. E. Arnold et al. hypothesised that listeners might expect a disfluent article to precede something new rather than something given in the display. In line with this, listeners demonstrated a bias toward the given object following fluent instructions; disfluent instructions, however, elicited a bias toward the new object. Crucially, the time course of listeners' fixations showed that this bias emerged 200 ms after onset of the temporarily ambiguous noun ("ca..."), almost as soon as the relevant acoustic input was available. J. E. Arnold et al. (2004) concluded that the fluency of a spoken utterance influences the initial stages of listeners' on-line processing, with a disfluency signalling information that is new to the discourse.

J. E. Arnold et al. (2004) later extended their findings to listeners' predictions of unfamiliar objects perceived as difficult to describe (J. E. Arnold, Kam, & Tanenhaus, 2007). Using a similar paradigm in which listeners followed recorded instructions to click on an object in a display, they showed that disfluent instructions ("Click on thee, uh...") biased listeners to look more at unfamiliar objects with no conventional name (e.g., Hebrew letters and engineering symbols) over everyday objects such as ice cream cones. In contrast, fluent instructions did not result in any particular bias toward familiar or unfamiliar objects. As with the earlier study, listeners' eye movements revealed that the disfluency effect emerged rapidly, around 200 ms after the point in the utterance at which the relevant linguistic information (the object description) began. J. E. Arnold et al. cite their findings across both studies as evidence that listeners interpret disfluencies as a signal of speaker difficulty, either due to describing discourse-new or unconventional objects. Importantly, their results show that these signals are brought to bear during the initial stages of comprehension, creating probabilistic expectations regarding the object they expect the speaker to refer to.

A parallel strand of event-related-potential (ERP) research has examined listeners' expectations in response to disfluency. These studies focus on the N400, an ERP component that provides a dependent measure of the ease of meaning processing. The N400, so named because it peaks around 400 ms post-onset of the relevant acoustic input, is known to reflect the semantic fit of a word with its surrounding context (Kutas & Hillyard, 1980, 1984). The N400 amplitude is inversely related to the degree of fit between the word and its context, with larger negative amplitudes observed for words that are more unexpected given the unfolding sentence or discourse (e.g., "He planted string beans in his *car*"). Corley, MacGregor, and Donaldson (2007) reasoned that if listeners' expectations regarding an upcoming noun are sensitive to the presence of a disfluency earlier in the utterance, this effect should be reflected in a difference in the N400 amplitude for disfluent compared to fluent utterances. In their study, listeners' ERP responses were measured as they heard sentences that ended with highly predictable or unpredictable words (e.g., "That drink's too hot; I have just burnt my *tongue/nails*"). Half of these included a disfluency just before the target word, in the form of an *uh* alongside a lengthening of the preceding word. As predicted, Corley et al. observed a characteristic N400 effect following target words that were unpredictable, reflecting listeners' difficulty in integrating lexical items that are unexpected given the preceding sentence context. Crucially, this N400 effect was greatly attenuated when the unpredictable word occurred after a disfluency, suggesting that the disfluency reduced listeners' difficulty integrating the word with its preceding context.

In a subsequent study, MacGregor, Corley, and Donaldson (2010) observed a similar effect with disfluent utterances that contained a silent pause. Using a similar paradigm, they demonstrated that the N400 effect for unpredictable compared to predictable words during fluent utterances was attenuated when the word was preceded by a silent pause. Taken together, results from the two studies suggest that disfluency, either in the form of a filled or a silent pause, can facilitate listeners' integration of unpredictable words during on-line comprehension. These conclusions parallel J. E. Arnold et al.'s findings that listeners' expectations about an upcoming word are rapidly affected by a preceding disfluency in the utterance.

Both the eye-tracking and the ERP studies thus establish that the manner of delivery of an utterance has consequences for listeners' on-line hypotheses about the speech: Disfluent utterances can create expectations about the upcoming message content or facilitate integration of an unexpected sentence ending. However, these studies focus on the effect of disfluency on listeners' literal interpretation of the unfolding message (in all cases, disfluency influences listeners' predictions about the semantic content of an upcoming item); considerably less work has explored how this aspect of manner may affect listeners' on-line pragmatic hypotheses beyond the surface word meaning. One reason for this prioritisation of the literal is due to traditional linguistic theory, which maintains an assumption that pragmatic interpretation is secondary to literal understanding, developing after listeners construct a literal utterance meaning (Grice, 1975). Under this account, any form of pragmatic comprehension is slow to take place, only occurring downstream at a later stage of processing (see section 2.3 for a detailed discussion).

Nevertheless, a small but growing literature is turning its attention to the role of manner of delivery in the ongoing process of pragmatic interpretation. These studies provide some evidence that a speaker's manner impacts a listener's on-line hypotheses surrounding the pragmatic meaning of an utterance. In particular, recent studies show that tonal composition and pitch contour can drive listeners' inferences about a contrastive meaning, by evoking an alternative set that is implicated but not explicitly mentioned (Dennison & Schafer, 2010; Esteve-Gibert, Portes, Schafer, Hemforth, & D'Imperio, 2016; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014).

(5) a. It LOOKS like a zebra
   b. ↳ It LOOKS like a zebra but it's actually not one

Building on earlier work which found that a verb-focus prosody on constructions such as in (5a) triggered the inference in (5b) (Kurumada et al., 2012), Kurumada et al. (2014) examined the time course with which these inferences are formed based on the prosodic contour of the utterance and the available visual context. Listeners in their study clicked on objects in a visual display as they heard recorded utterances of the construction "It looks like an X". The utterances were produced with either a canonical noun-focus contour with nuclear pitch accent on the noun, or a verb-focus contour with a contrastive accent on "looks" and a rising boundary tone. The displays included a contrast pair of objects comprising a prototypical and a non-prototypical member (e.g., a zebra vs. an okapi—a zebra-like animal), and two distractor objects. In line with results from their off-line study, Kurumada et al. found that utterances with verb-focus prosody elicited more early fixations to the non-prototypical member object (the okapi), demonstrating that listeners make use of prosody incrementally to formulate pragmatic expectations about an implied contrastive meaning. Notably, in conditions where the distractor images consisted of two unrelated objects (e.g., a pig and a ladybug), listeners' fixation biases toward the non-prototypical target following a verb-focus prosodic contour emerged earlier compared to when the distractors comprised a second pair of contrast objects (e.g., a typical ladybug and an atypical ladybug). In the former condition, listeners made anticipatory eye movements to the non-prototypical target following a contrastive accent on "looks" and prior to noun-onset; in contrast, in the latter condition this fixation bias emerged about 200 ms after noun-onset, the point in the utterance at which the visual context allowed listeners to uniquely identify one of the objects in the target contrast pair. This finding suggests that listeners' pragmatic inferences based on prosodic information are context-dependent: Listeners take into account information available in both the manner of delivery and the visual context to shape their overall meaning of the utterance.

Van Berkum, Van den Brink, Tesink, Kos, and Hagoort (2008) investigated the effect of manner of delivery on a different form of pragmatic interpretation: that relating to a listener's inferences about the identity of the speaker. They recorded listeners' ERP responses as they heard recordings of sentences produced by a mix of speakers in Dutch. Critically, some utterances were characterised by an inconsistency in the form of a word that defied probabilistic social stereotypes associated with the speaker's sex, age or social-economic status; for example:

(6) a. "If only I looked like *Britney Spears* in her latest video" in a male voice (gender inconsistency)
   b. "Every evening I drink some *wine* before I go to sleep" in a child's voice (age inconsistency)
   c. "I have a large *tattoo* on my back" in an upper-class accent (socio-economic status inconsistency)

In addition, their stimuli included sentences that featured a standard semantic anomaly typically evocative of an N400 response (e.g., "The earth revolves around the *trouble* in a year.")

Based on evidence from speech perception that listeners extract information about a speaker's identity, such as their age or gender, based on the acoustic signal (Strand, 1999), Van Berkum et al. hypothesised that listeners would bring to bear the social stereotypes associated with these aspects of speaker identity during their comprehension of the linguistic message. They further reasoned that if these voice-based inferences about speaker identity are recruited by the same early sense-making processes responsible for word-meaning construction, speaker-inconsistencies in an utterance should result in the same N400 effect that semantic anomalies elicit. Consistent with this, their results showed a small but clear N400 effect from the point in the utterance at which inconsistency between the unfolding message and listeners' voice-based inferences about the speaker emerged. These results demonstrate listeners' use of voice-based cues to form pragmatic inferences about a speaker's identity, as well as the speed with which these cues are brought to bear on the construction of meaning based on the message.

The results of Kurumada et al. (2014) and Van Berkum et al. (2008) show that manner of delivery can rapidly constrain a listener's on-line pragmatic interpretations, whether in the form of hypotheses surrounding an alternative (implicated) meaning or inferences about a speaker's identity and its associated social stereotypes. These findings are difficult to reconcile with the traditional view that pragmatic interpretation is slow and costly for the comprehension system (e.g., Bott & Noveck, 2004). Rather, Van Berkum et al. interpret their results within a one-step model of comprehension, in which pragmatic information about the speaker and context is integral to the construction of meaning from the earliest stages of understanding (cf. Grodner, Klein, Carbary, & Tanenhaus, 2010; Nieuwland & Van Berkum, 2006). In the next section, we introduce two contrasting models of comprehension—the *standard pragmatic model* and the *one-step model*, and discuss the theoretical implications of each for the time course of pragmatic comprehension.

2.3 The time course of pragmatic processing

The recent interest in the effect of manner on listeners' on-line comprehension is part of a growing focus in psycholinguistic research on examining the mechanisms underlying real-time language comprehension. Over the past few decades, an increasing number of studies have turned their attention to the ongoing process of comprehension, using fine-grained, time-sensitive measures such as eye-tracking and electroencephalographic (EEG) recordings to study how continuous speech is interpreted from moment to moment as the utterance evolves. However, while a good deal of research has focused on examining the time course of semantic interpretation (e.g., Huettig & Altmann, 2005; Kamide, Altmann, & Haywood, 2003; Kim & Osterhout, 2005; Kutas & Hillyard, 1984; Yee & Sedivy, 2006), considerably less attention has been paid to the pragmatic aspects of on-line processing. A large part of this is a result of traditional linguistic theory, which emphasises a distinction between the literal and the pragmatic interpretations of an utterance. This distinction is held by the *standard pragmatic model* of comprehension (Grice, 1975; Searle, 1969). Crucially, the model postulates a prioritisation of the literal, with pragmatic meaning viewed as more difficult to access than its semantic counterpart. Consequently, this has evolved into an assumption that pragmatic understanding is relegated to a later, post-literal stage of processing, demanding additional time and effort on top of its literal counterpart.

However, recent studies have questioned the validity of this claim. A growing body of research now challenges the literal-first stance held by traditional linguists. Studies investigating the relative time course of semantic and pragmatic processing demonstrate that contrary to predictions by the standard pragmatic model, listeners can often derive pragmatic interpretations as quickly as literal meanings. Some proponents of this view argue instead for an interactive, *one-step model* of comprehension, in which pragmatic knowledge is brought to bear immediately by the same early fast-acting system responsible for deriving the literal meaning of an utterance (Grodner et al., 2010; Nieuwland & Van Berkum, 2006; Van Berkum et al., 2008). The standard pragmatic model and the one-step model thus make competing claims about the time course of pragmatic interpretation: The former maintains that pragmatic interpretation occurs at a late, post-literal stage of comprehension, while the latter asserts that pragmatic meaning can be derived from the earliest moments of comprehension. This section provides an overview of the two models, looking at how each proposes the comprehension system integrates pragmatic information in the process of meaning construction.

2.3.1 The standard pragmatic model

Fundamental to the standard pragmatic model of comprehension is a prioritisation of the literal meaning of an utterance. Under this account, listeners must first compute a local, context-independent surface meaning before drawing on any pragmatic information to infer a contextually-appropriate non-literal interpretation. This position has roots in the Gricean tradition of conversational cooperation, which assumes that speakers are bound to the conversational maxims that govern the Cooperative Principle (Grice, 1975). As noted previously, Grice observed that maxims can be flouted to convey meaning over and above what is literally said. Accordingly, from the hearer's perspective such cases demand a sequential process: A recognition that a violation of a maxim has occurred, subsequent rejection of the literal meaning, then an inferential process through which the actual intended meaning is derived (cf. Giora, 2003; Handl, 2011; see also Searle, 1979 for a proposal of a similar mechanism underlying metaphor comprehension). The underlying premise therefore is that pragmatic meaning represents a deviation from the "literal norm", requiring interpretation in a two-step fashion. The first step is literal and non-optional, by which a context-free, compositional meaning is activated by combining the core word meanings; the second step is non-literal and optional, by which the literal meaning is evaluated against the wider communicative context, tagged as inappropriate, and replaced with an alternative, contextually appropriate meaning. This additional step in processing has led many researchers to refer to the standard pragmatic model as a *two-step* model of processing (e.g., Hagoort & van Berkum, 2007; Nieuwland & Van Berkum, 2006; Tomlinson, Bailey, & Bott, 2013).

One notable source of evidence for mandatory literal-first computation comes from studies on the interpretation of the quantifier *some*. This line of research provides a useful testing ground for investigating the relative unfolding of literal and pragmatic comprehension, since the quantifier *some* is locally ambiguous between a literal and a pragmatic interpretation. Several studies demonstrate that the pragmatic interpretation, *some but not all*, is significantly delayed relative to its literal *some and possibly all* counterpart (Bott & Noveck, 2004; Huang & Snedeker, 2009a, 2009b, 2011; Noveck & Posada, 2003; Tomlinson et al., 2013). Bott and Noveck (2004) asked participants to perform a truth-value judgement task on categorical sentences containing the quantifier *some*:

(7) Some elephants are mammals

Participants read sentences like (7) and were tasked with providing a True/False judgement on each sentence. In one session, participants were instructed to interpret *some* literally, while in another they were instructed to interpret it pragmatically. Bott and Noveck measured participants' proportions of correct responses and response times, and found that participants were less accurate and took significantly longer to respond when under instructions to interpret *some* pragmatically. A subsequent experiment which imposed no restrictions on how to interpret *some* showed that participants who intuitively responded with a pragmatic *some but not all* interpretation took longer than those who responded with a literal *some and possibly all* interpretation. Bott and Noveck's findings are consistent with those of an earlier study by Noveck and Posada (2003), which showed that listeners took longer to respond pragmatically to underinformative sentences such as "Some dogs have ears" than they did to respond literally. More broadly, they also complement findings from several studies on figurative language processing. These studies indicate a longer processing time associated with the pragmatic interpretation of various forms of figurative language like irony (Dews & Winner, 1999; Giora & Fein, 1999; Schwoebel, Dews, Winner, & Srinivas, 2000), metaphors (Blank, 1988; Gregory & Mergler, 1990; Janus & Bever, 1985), proverbs (Honeck, Welge, & Temple, 1998; Temple & Honeck, 1999) and idiomatic expressions (Gibbs, Nayak, & Cutting, 1989), compared to equivalent literal speech.

The above-mentioned studies suggest a processing cost associated with accessing the pragmatic meaning of a sentence, lending weight to the claim that pragmatic interpretation is slow and computationally-intensive. Nevertheless, these findings do not directly signify a literal advantage. Although a comparison of response times indicates that accessing the pragmatic interpretation takes longer, this difference does not explicitly demonstrate that the delay is due to having to access a literal meaning first.

Time-sensitive measures of processing provide a more direct means of addressing the question of when listeners access literal and pragmatic interpretations. One such study was conducted by Tomlinson et al. (2013), who tracked participants' mouse movements as they evaluated sentences like (7). The sentences were presented word-by-word on a screen, and participants' mouse coordinates were recorded as they selected between a True/False response following each sentence. An analysis of participants' average mouse trajectories revealed that when responding with a literal (i.e. True) interpretation, their mouse paths exhibited a direct arc toward the True response; in contrast, when responding with a pragmatic (i.e. False) interpretation, participants' mouse trajectories displayed a tendency to deviate toward the True response before changing direction mid-flight to select the False target. A similar pattern was observed whether participants underwent an initial training phase with feedback to encourage either a literal or a pragmatic interpretation (Experiment 1), or were allowed to respond with their most natural interpretation of *some* (Experiment 2). Tomlinson et al. interpreted their results within a two-step model of scalar comprehension, where a literal interpretation is first computed, before listeners apply the Cooperative Principle and general reasoning to evaluate the validity of the implied claim that some, *but not all,* elephants are mammals.

One potential criticism that Tomlinson et al.'s study faces (which similarly applies to Bott & Noveck, 2004) relates to the naturalness of the sentences being judged. As Geurts (2010) points out, the question of whether some elephants are mammals is not one that typically arises in everyday life. Notably, the pragmatic interpretation that listeners have to derive in this context is one that conflicts with common knowledge (cf. Magri, 2009). This raises the concern of the degree to which the artificial nature of the judgements may have driven participants' reasoning processes. A second limitation is that participants were presented with sentences in text rather than spoken form. Although mouse movements provide a fine-grained, time sensitive measure of the ongoing process of comprehension, the comprehension of written and spoken language likely evoke distinct processes (Booth et al., 2002), with reading times influenced by a number of factors that invoke syntactic and pragmatic processes (Murphy, 1985). A related concern is the possible response advantage associated with literal interpretations, for which participants responded with a "True" rather than "False" judgement (although see Bott & Noveck, 2004, Experiment 2 for evidence to counter this argument).

Huang and Snedeker (2011) overcome some of these limitations faced by the above-mentioned studies by employing a visual world eye-tracking paradigm, in which participants map spoken utterances to images in a visual reference scene (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; see section 2.4.1 for a detailed introduction). This approach removes overt task demands that may accompany categorical judgement tasks (cf. Huang & Snedeker, 2009b). In their study, listeners viewed displays featuring four characters, two of each gender, with two sets of items distributed amongst them. The names of the two items were compound nouns whose onsets were temporarily ambiguous (e.g., ice cream cones and ice cream sandwiches). Listeners' eye movements were tracked as they followed audio instructions such as "Point to the girl that has *some/two/all/three* of the ice cream sandwiches". Of particular interest were *some* trials, on which listeners could map the instruction to two characters—a girl with a subset of the ice cream sandwiches (the target) or a girl with all of the ice cream cones (a distractor)—prior to disambiguation. Huang and Snedeker observed that following onset of the quantifier *some*, listeners were initially equally likely to fixate both girls, with looks to the target not reliably exceeding chance until 1,100 ms post-onset. In contrast, target fixations rose above chance within 200–300 ms post-quantifier onset on *all/two/three* trials. This suggests that the pragmatic meaning of *some* was not available during early processing to rule out a literal interpretation, with listeners initially interpreting *some* as compatible with *all*.

A further analysis based on listeners' saccades showed that in the 800 ms time window post-*some* onset, listeners were more likely to switch from looking at the distractor (girl with all of the ice cream cones) to the target (girl with a subset of the ice cream sandwiches) than from the target to the distractor. This finding suggests an extended period where *some* was interpreted literally before the pragmatic inference was derived around 800 ms after onset of the relevant acoustic input (cf. Huang & Snedeker, 2009b, Experiment 3). This time course of events is also highly consistent with the pattern observed in an earlier study, whereby listeners asked to "Point to the girl with some of the socks" displayed equal fixations to the literal (girl with three of three soccer balls) and the pragmatic (girl with two of four socks) targets prior to disambiguation (Huang & Snedeker, 2009a).

Taken together, these findings paint a convincing picture: Pragmatic comprehension incurs a processing cost, with this cost arising from having to first compute the literal meaning before a pragmatic interpretation is derived. This perspective, however, has not gone unchallenged. In particular, Grodner et al. (2010)

highlight limitations to the Huang and Snedeker studies, such as their frequent use of utterances featuring exact numbers (two/three), which may have reduced the acceptability of a pragmatic interpretation of *some* (cf. Degen, Reeder, Carbary, & Tanenhaus, 2009). Using a modified design and materials, they demonstrate that with appropriate contextual support, processing delays associated with pragmatic-*some* disappear, and argue instead for a model that supports the immediate integration of pragmatics from the early stages of comprehension. In the next section, we introduce an example of such a model, the one-step model of comprehension.

2.3.2 The one-step model

In direct contrast to the two-step processing mechanism of the standard pragmatic model, the one-step model posits that various sources of information are taken into account simultaneously in the process of meaning construction. This may include the local contribution of semantic information at the word-level, but also input from the wider communicative context such as knowledge about a speaker (Van Berkum et al., 2008), the prior discourse (Nieuwland & Van Berkum, 2006; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003), real-world knowledge (Hagoort et al., 2004; Hagoort & van Berkum, 2007), a speaker's manner (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Kelly, Kravitz, & Hopkins, 2004; Kurumada et al., 2014) or the visual context (Willems, Özyürek, & Hagoort, 2008). Under this account, all available sources of information are integrated immediately to constrain a listeners' interpretation of the unfolding linguistic message (the *immediacy assumption*; cf. Just & Carpenter, 1980). Consequently, the one-step model does not draw a sharp temporal distinction between the literal and pragmatic aspects of comprehension. Rather, pragmatic information is allowed to influence interpretation immediately, without first giving precedence to a context-free, literal interpretation.

One of the strongest arguments for the immediate integration of pragmatics comes from Grodner et al. (2010), who provide direct evidence to counter Huang and Snedeker's (2009a; 2009b; 2011) claim that a literal interpretation of *some* has to be computed before the pragmatic meaning is derived. Like Huang and Snedeker, Grodner et al. investigated whether the pragmatic interpretation of *some* is universally delayed, using a modified design and materials to reduce the effects they attributed to Huang and Snedeker's results.

Building on the logic that using an exact number is more natural than *some* for small sets (e.g., 2,3; Degen & Tanenhaus, 2010, 2014), Grodner et al. reasoned that the pragmatic delay observed by Huang and Snedeker could have resulted from the inclusion of instructions such as "Point to the girl with *two* of the...", thereby reducing the felicity of instructions on *some*-trials, where the target had two of the full set of objects. Additionally, Grodner et al. observed that *some* alone is not unambiguously associated with the meaning *not all* (cf. Degen et al., 2009), and convergence on pragmatic *some* in Huang and Snedeker's studies might have been delayed until after processing of the partitive construction *of*. In response to these concerns, Grodner et al. designed their study to eliminate instructions with exact quantities, and replaced *some of* with the shortened form *summa* (*all of* and *none of* controls were correspondingly shortened to *alla* and *nunna*). They used displays that featured three boys and three girls, with three sets of objects distributed amongst them. Two of the objects were phonological cohorts (e.g., balls and balloons), with these distributed such that one individual had a subset of the items of one cohort (e.g., two of four balls) while another same-gender individual had all the items of the other cohort (e.g., all four balloons). Listeners followed instructions to "Click on the girl who has *summa/alla/nunna* the...". Grodner et al. compared fixations to the target individual and gender competitor on each display, and found that convergence on the appropriate target was just as fast for *summa* utterances as for *alla* ones. These effects emerged within 200–300 ms after onset of the quantifier—the earliest point within the utterance at which listeners could identify a referent based on the available phonological information, and 800 ms faster than the time course observed by Huang and Snedeker for the pragmatic interpretation of *some*. In other words, Grodner et al. found no support for a literal interpretation of *some* to mean *all* prior to its pragmatic interpretation. Grodner et al.'s findings are inconsistent with Huang and Snedeker's, and difficult to reconcile with the hypothesis that literal meaning is necessarily computed prior to pragmatic interpretation. Instead, their results support a one-step model of processing, where pragmatic information is taken into account to shape interpretation from the earliest moments of comprehension.

Grodner et al. further suggest that the increased processing time for pragmatic-*some* observed by earlier studies is not inherent to the actual generation of the pragmatic inference. Rather, they attribute this to difficulty in integrating its meaning with the available information in the context, such as considerations about alternative forms that the speaker had previously used (e.g., "two" instead of "some" to describe an individual with two of four items). Based on their findings, they argue that with adequate contextual support, processing delays associated with pragmatic interpretation disappear. This perspective reinforces the immediacy of contextual information in constraining interpretation as proposed by the one-step model.

The notion of contextual support is also implicated by a number of earlier studies on non-literal language processing. These studies demonstrate that the processing disadvantage for non-literal language disappears when such expressions are well-supported by context (Gerring & Healy, 1983; Gibbs & Gerrig, 1989; Gildea & Glucksberg, 1983; Glucksberg, Gildea, & Bookin, 1982; Keysar, 1989; Pfaff, Gibbs, & Johnson, 1997; see Gibbs & Colston, 2006 for a review). Evidence from reading time studies, for instance, shows that metaphoric expressions are read as quickly as their literal equivalents when embedded in longer paragraphs that provide an adequate context for interpretation (Inhoff, Lima, & Carroll, 1984; Ortony, Reynolds, & Arter, 1978). Similar effects have been observed with the comprehension of sarcasm: Gibbs (1986a, 1986b) found that the context in which a sarcastic statement was presented facilitated understanding such that the sarcastic meaning took no longer, and in some cases less time, to process than literal uses of the same expression. The overarching message behind these findings is that the longer processing time associated with non-literal language may not be due to additional time needed to compute and reject a literal meaning, but to difficulty integrating the pragmatic meaning within its context (cf. Gibbs, 2002; Gibbs & Colston, 2006). Although these studies do not directly interpret their results within a one-step model of processing, they nevertheless provide support for the underlying notion that with sufficient contextual information, pragmatic interpretation can occur immediately, without giving temporal precedence to literal comprehension.

The above results from non-literal language processing, however, bear the same constraints that limit studies employing off-line measures of response (e.g., Bott & Noveck, 2004). That is, verification judgements and response time measures only provide an indirect indication of a listener's moment-to-moment interpretation of the unfolding linguistic input. In other words, although their results challenge the literal advantage maintained by the standard pragmatic model, they are limited in what they tell us about when and how the pragmatics of an exchange is integrated by a listener during the process of meaning construction.

A recent line of ERP research addresses these questions by investigating the neural integration of various forms of contextual information on a listener's moment-to-moment interpretation. The research logic of these studies is based on two established characteristics of the N400 component: a) its sensitivity to various types of unexpectancy (Ganis, Kutas, & Sereno, 1996; Kutas, Neville, & Holcomb, 1987; Sitnikova, Kuperberg, & Holcomb, 2003), and b) the temporal localisation of effects within the unfolding stimulus (a 200–600 ms window, typically peaking at around 400 ms post-stimulus onset; Kutas & Federmeier, 2011). Building on these two facts, researchers have begun to explore the interplay between local, semantic features and global, contextual information on incremental language processing. These studies provide temporal evidence to show that, in line with predictions by a one-step model, information beyond the local contribution of core word meanings is recruited from the earliest stages to constrain a listener's interpretation (e.g., Dick et al., 2009; Hagoort et al., 2004; Nieuwland & Van Berkum, 2006; Van Berkum et al., 2008; see Van Berkum, 2012 for a review).

Information from the wider discourse context, for example, has an immediate impact on comprehension. Nieuwland and Van Berkum (2006) showed that the N400 effect typically associated with a predicate that is anomalous out of context (e.g., "The peanut was *in love*") can be eliminated by embedding the sentence within a supportive discourse context, such as a cartoon story about an animate peanut and his girlfriend. Notably, the reverse was observed for semantically acceptable but contextually inappropriate sentences (e.g., "The peanut was *salted*"): When presented within the same story about an animate peanut, these canonical, but story-irrelevant predicates elicited a clear N400 effect. These findings strongly suggest that the wider discourse context can overrule meaning violations at the local, word-level, lending support to a model in which local semantics and global context immediately interact to influence interpretation. Listeners' early interpretations show similar sensitivity to their knowledge of the world beyond the discourse: Sentences such as "Dutch trains are *white* and very crowded" (at odds with the well-known fact among Dutch people that Dutch trains are *yellow*) elicit an N400 difference comparable with semantic violations such as "Dutch trains are *sour* and very crowded" (Hagoort et al., 2004). Voice-based inferences about a speaker's identity have also been shown to modulate N400 activity, with N400 effects observed for linguistic content that mismatches what a speaker is likely to say given their gender, age or socio-economic status (Van Berkum et al., 2008). These studies provide converging evidence of the comprehension system's ability to rapidly access and integrate information beyond the lexical-semantic material. The time course of the ERP data further suggest that these global, contextual cues are brought to bear by the same early brain system responsible for deriving a surface word meaning. While these findings are difficult to reconcile with the literal-first assumption held by the standard pragmatic model, they support an interactive, one-step model by which the construction of meaning is coloured by the pragmatics of communication from the earliest moments of comprehension.

In summary, the standard pragmatic model and the one-step model represent conflicting perspectives on the time course of pragmatic comprehension. The standard pragmatic model maintains that listeners must analyse the literal content of utterances before pragmatic enrichment can occur; in contrast, the one-step model proposes that pragmatic information is integrated from the earliest stages of comprehension, without having to give temporal precedence to a literal interpretation. This debate is still ongoing, although there is growing evidence in favour of rapid pragmatic interpretation, particularly when adequate contextual support is provided. We additionally note that findings from the studies we present in this thesis suggest that, contrary to predictions by the standard pragmatic model, listeners take into rapid account a speaker's manner of delivery to influence their on-line pragmatic hypotheses about the speaker's utterance. This is evident from listeners' fixations as well as their mouse movements that accompany the unfolding linguistic input. In the next section, we discuss both of these methodologies in greater detail. With reference to studies within the current literature, we demonstrate their validity as an effective measure of the cognitive processes underlying language comprehension.

2.4 Analysing linguistic processing in real-time

The rapidly-expanding body of research on real-time linguistic processing has been made possible in part due to methodological advancements in the field. One method that has undergone considerable development in the last few decades is the use of eye movements to study cognitive behaviour. Although eye-tracking technology has existed for over a century, it was not until the mid-1970s that researchers began to tap more deeply into its potential. This movement arose partly as a result of marked improvements to eye-tracking systems, which increased the accuracy and precision of the measures obtained, and reduced the overall invasiveness of the procedure on participants (Jacob & Karn, 2003; Young & Sheena, 1975). This era of technological advancement also enabled eye-trackers to be interfaced with laboratory computers, allowing researchers to gather large quantities of data via a range of innovative techniques (Rayner, 1998).

Eye-tracking has since proven to be a valuable tool in studying a range of information processing tasks such as reading, visual search and scene perception (see e.g., Clifton, Staub, & Rayner, 2007 for a review). More recently, a growing interest has developed in the use of eye-tracking to investigate spoken language processing. Research in this area expanded following the emergence of the visual world paradigm (Cooper, 1974; Tanenhaus et al., 1995), which demonstrated that a listener's gaze behaviour is sensitive to unfolding linguistic input. Using this method, researchers are able to analyse listeners' gaze patterns in response to speech over time, yielding a real-time measure of linguistic processing without having to interrupt the speech stream.

2.4.1 Eye-tracking and the visual world paradigm

The visual world paradigm was pioneered by Cooper (1974) and further developed by Tanenhaus et al. (1995). The basic set-up of the paradigm involves monitoring participants' eye movements as they attend to a visual display comprising some objects presented alongside spoken language. Cooper's study first showed that individuals' eye movements were drawn to objects on a display that were referenced by accompanying speech. For instance, on hearing the word "dog" within a narrative, participants were more likely to direct their gaze toward a picture of a dog compared to unrelated pictures in the display. Cooper's findings had a fairly limited impact on the psycholinguistic community, perhaps due to the relatively cumbersome nature of eye-tracking techniques at the time (cf. Huettig et al., 2011). His findings were later extended by Tanenhaus et al. (1995), who used a similar method to explore listeners' gaze behaviour following spoken instructions. They observed that instructions such as "Touch the starred yellow square" elicited more saccadic eye movements following the word "starred" toward the only starred block out of four shapes on the display, demonstrating that listeners' eye movements are tightly linked to the unfolding speech and visual context. Following their study, a growing body of research began to exploit this systematic relationship between the linguistic and visual processing systems, using language-mediated eye movements to explore the temporal dynamics of the processes underlying comprehension (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Huettig & Altmann, 2005; Kamide et al., 2003; see Huettig et al., 2011 for a review).

In a typical visual world paradigm, listeners are presented with a number of objects in a display, accompanied by spoken language. The objects typically include one or more referents mentioned in the sentence(s) alongside other distractors. For example, in Huettig and Altmann (2005), participants were shown a display depicting a piano, a goat, a carrot, and a hammer as they heard the sentence "Eventually, the man agreed hesitantly, but then he looked at the piano and appreciated that it was beautiful". The linguistic input is understood to guide the listener's attention, while the scene constrains the domain of referents in the visual field upon which the speech can be mapped. The temporal sensitivity of eye movements allows researchers to time-lock analyses to a particular point of interest, normally the start of a critical word or phrase in the acoustic stimuli. For instance, Huettig and Altmann observed more looks directed toward the piano following its onset, compared to looks to the distractor objects. These effects were triggered rapidly, with fixations converging on the piano within 200 ms after onset, demonstrating listeners' eye movement sensitivity to objects in their visual field in response to the speech.

By manipulating the relationship between the objects and the speech, researchers are able to draw conclusions about how and when listeners formulate hypotheses based on the unfolding linguistic input. For example, Huettig and McQueen (2007) used sentences like "Eventually she looked at the *beaker* that was in front of her", alongside displays that contained a phonological (*beaver*), a shape (*bobbin*) and a semantic (*fork*) competitor, as well as a distractor (*umbrella*). They found that fixations to the phonological competitor preceded fixations to the shape and semantic competitors, but nonetheless looks to all three overlapped in time. Their findings suggest that spoken word recognition is a cascaded process, where processing at one level does not have to be completed before information from other levels is retrieved.

Since its advent, the visual world paradigm has been used to study a range of linguistic phenomena, such as phonological processing (Allopenna et al., 1998; Desroches, Joanisse, & Robertson, 2006), semantic prediction (J. E. Arnold et al., 2007, 2004), ambiguity resolution (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995), visual-shape pre-activation (Dahan & Tanenhaus, 2005), and even dyadic interaction (Nadig & Sedivy, 2002). Notably, studies have recently begun to explore its potential in investigating on-line pragmatic comprehension (Grodner et al., 2010; Huang & Snedeker, 2011; Kurumada et al., 2014). The visual world paradigm also offers a promising tool for researchers investigating processes that might be unique to spoken language comprehension, such as comprehension of a speaker's manner of delivery. Bailey and Ferreira (2007) provide an example of how a speaker's disfluencies may affect a listener's on-line parse of an utterance, but not their final interpretation. Based on previous findings that disfluencies are more likely to precede a complex syntactic constituent (Clark & Wasow, 1998), Bailey and Ferreira examined whether disfluencies can correspondingly cue the upcoming structure of a syntactically ambiguous utterance. Listeners in their study followed spoken instructions in the form of a garden-path sentence, with disfluencies in different locations to bias either a modified-theme interpretation (e.g., "Put the *uh uh* apple on the towel in the box)" or a modified-goal interpretation (e.g., "Put the apple on the *uh uh* towel in the box"). They found that the location of the disfluency had no effect on listeners' final interpretation of the instruction; however, the eye movement data suggested a trend for more looks to be directed to the towel in a box when the disfluency occurred in the modified-goal biasing location than in the modified-theme biasing location. Their results reflect the perceptual relevance of spoken cues, such as a speaker's manner of delivery, on the moment-to-moment processing of a linguistic utterance (cf. J. E. Arnold et al., 2007, 2004; Barr, 2001; Barr & Seyfeddinipur, 2010). Although the final parse of the utterance may not be affected, disfluency can impact the ongoing process of comprehension; accordingly, tracking listeners' eye movements within a visual world paradigm can yield a closely time-locked and fine-grained measure of these ongoing processes.

However, this method is not without is limitations. One potential limitation relates to how visual world data is typically analysed. Data analyses in visual world studies focus on how likely listeners are to look at specific regions at different points in time. The visual displays allow researchers to pre-define regions of interest, normally the areas occupied by individual objects in the display. The eye movements are typically analysed in terms of fixations and saccades (e.g., Altmann, 2004): Fixations to a region of interest, time-locked to the onset of a critical word or phrase, are taken as an indicator of activation of the corresponding word or phrase at that point (Allopenna et al., 1998). To determine how the pattern of fixations changes over time, most researchers break time up into a series of consecutive windows (e.g., 100 ms), performing separate t-tests or analyses of variance (ANOVAs) on each window to compare whether two regions differ in their likelihood of being fixated at that point (e.g., J. E. Arnold et al., 2004; Huang & Snedeker, 2011; Nadig & Sedivy, 2002). Accordingly, the dependent variable is typically the proportion of fixations to a region of interest during each time window, aggregated over time and over multiple experimental trials.

However, this approach may be problematic on several fronts. The transformation of the categorical variable of fixation region into a continuous one by calculating the proportion of fixations to that region violates the ANOVA assumptions of an unbounded range and normal distribution of errors. The transformation of time, inherently continuous in nature, into a categorical variable by collapsing it into discrete windows can also obscure theoretically relevant data patterns (cf. Barr, 2008). Furthermore, the contribution of multiple data samples from each participant due to repeated sampling by the eye-tracker violates the assumption of independence of observations, as the observations associated with a given trial for one participant would be more correlated with one another than the observations collected on another trial. Barr (2008) outlines a solution to these problems by proposing an analysis framework using multilevel logistic regression, in conjunction with an empirical logit transformation that maps the proportional dependent variable onto a scale that is unbounded and symmetric around zero. This accommodates the categorical variable of gaze location, and allows the researcher to incorporate time as a continuous predictor in their model. The regression framework also resolves the problem of non-independence by modelling random effects corresponding to the sampling clusters in the design (cf. Raudenbush & Bryk, 2002). With these considerations in mind, we apply Barr's framework to our analyses of eye-tracking data in this thesis.

A second limitation however, relates to the nature of eye-tracking data itself. The neural programming of the human visual system means that eye movements are largely ballistic, comprising a series of fixations (periods of relative stability) and saccades (rapid movements between fixations; Salvucci & Goldberg, 2000). From the point of view of a visual world experiment, this results in fixations that are inherently discrete, either falling on an object of interest or not, on any given trial. Consequently, analyses tend to rely on the frequency of discrete fixations to objects within a display. This measure may, however, fail to fully capture the graded nature of perceptual-cognitive processing over time, which evolves in a continuous, dynamic trajectory rather than a series of discrete steps (cf.

Spivey & Dale, 2006). For example, Farmer, Cargill, Hindy, Dale, and Spivey (2007) point out that eye-tracking results fall short of distinguishing between discrete stage-based and graded constraint-based accounts of syntactic ambiguity resolution, both of which can result in a similar pattern of eye movements for garden-path sentences observed across trials. In response to this, some researchers have begun to explore dynamic motor movements as a means of accessing the temporal continuity of real-time linguistic processing. One example of this is based on listeners' mouse movements in response to speech.

2.4.2 Mouse-tracking

The use of computer mouse-tracking to assess real-time processing is a relatively recent development in the field of language research. This method involves recording the streaming x, y coordinates of the mouse cursor as participants move the mouse to select between responses on a screen. The output of this process is a continuous stream of motor measures that reflect the evolving mental trajectory, allowing the researcher to draw conclusions about the ongoing dynamics of comprehension (e.g., Spivey, Grosjean, & Knoblich, 2005).

The ubiquity of a computer and mouse environment allows mouse-tracking to be easily integrated within the visual world paradigm. As with eye movements, mouse movements can be time-locked to a particular point within a speech stream. By analysing these time-locked mouse trajectories in relation to the objects in a display, researchers can gain a dynamic picture of listeners' tentative commitment to alternative responses (Spivey & Dale, 2006). For example, Spivey et al. (2005) had listeners follow spoken instructions such as "Click the candle" whilst they viewed displays depicting a candle and a distractor object. They found that listeners' mouse trajectories displayed a marked attraction toward the distractor when its name shared a phonological onset with the target (e.g., *candy*) compared to when it did not (e.g., *pickle*). These findings provide evidence for the continuous uptake of acoustic-phonetic input during spoken word recognition, where partially active lexical representations may compete prior to completion of the word (cf. Allopenna et al., 1998).

The continuous trajectory provided by mouse movements additionally affords the investigation of comprehension processes that may reflect a temporally continuous mechanism rather than a stage-by-stage system. One such study was conducted by Farmer et al. (2007), who investigated listeners' processing of syntactically ambiguous garden-path sentences such as:

(8) Put the apple on the towel in the box

Farmer et al. noted that while earlier eye-tracking studies (e.g., Spivey et al., 2002; Tanenhaus et al., 1995) interpreted their results within a constraint-based account where all possible syntactic representations may simultaneously compete for attention, their data were equally compatible with a discrete stage-based explanation. This is because the pattern of fixations across all experimental trials could simply reflect a cumulation of discrete interpretations on individual trials, on which a listener either fixates the object associated with the garden-path interpretation or does not. Farmer et al. set out to distinguish between these two accounts by using a mouse-tracking paradigm in which participants manipulated objects on a screen following instructions such as (8). They found that when moving, for instance, an apple on a towel, listeners' mouse trajectories exhibited a spatial attraction toward the object associated with the incorrect parse (an empty towel) before reaching the correct destination (the box). These findings provide support for a constraint-based account where alternative syntactic representations may be simultaneously active as the parser attempts to disambiguate between them. Crucially, Farmer et al. were able to demonstrate this effect in the trajectory curvatures on a trial-by-trial basis, excluding a discrete stage-based interpretation that previous eye-tracking results were unable to rule out.

More recently, mouse-tracking has also proven an effective means of investigating the on-line comprehension of a speaker's manner of delivery. Barr and

Seyfeddinipur (2010) tracked participants' mouse movements as they identified abstract images on two-object displays based on a speaker's description of a target image. They varied the speaker's manner with an *um* or incidental noise (a cough or sniffle) before the description. Based on previous eye-tracking results to show that disfluencies shift listeners' expectations toward difficult-to-name objects (J. E. Arnold et al., 2007; Barr, 2001), Barr and Seyfeddinipur hypothesised that disfluencies might bias listeners toward referents that are new for a speaker. Consistent with this, they observed that listeners moved the mouse more toward targets when the descriptions were preceded by *ums* compared to noise. Barr and Seyfeddinipur further demonstrated that listeners' mouse movements following disfluency as a cue to difficulty are sensitive to speaker effects: This expectation varied depending on whether the description was produced by a speaker who was new to both images, or a speaker who was familiar with one of the images from a previous trial. These effects emerged over a time window spanning the onset of the disfluency to the onset of the description, demonstrating the rapidity with which these manner-based cues influenced listeners' language-mediated mouse movements.

These studies thus provide evidence for a close-knit relationship between the visual, cognitive and motor systems, highlighting the promise of mouse-tracking as an effective tool to study the continuous dynamics of real-time language processing. However, we note that this should not be taken as indication of any theoretical or functional advantage of mouse-tracking in comparison to eye-tracking. As Spivey et al. (2005) acknowledge, the two have their compensatory strengths and weaknesses: Eye movements are highly temporally sensitive, while mouse movements are highly spatially sensitive. Studies that have replicated existing eye-tracking paradigms using a mouse-tracking methodology observe that these motor movements tend to be launched slightly later than an initial eye movement (e.g., Farmer et al., 2007); what they offer in exchange is a smoother and more continuous index of the mental trajectory, allowing researchers to glean information about the interplay of graded mental states underlying comprehension. It is with this in mind that we employ both methods concurrently in this thesis to obtain a more informed picture of how a speaker's manner of delivery influences a listeners' pragmatic interpretation.

2.5 The current thesis

Thus far, we have established that the manner of delivery of an utterance impacts a listener's comprehension. Specifically, disfluencies influence listeners' off-line pragmatic judgements (Brennan & Williams, 1995), as well as their on-line expectations surrounding the upcoming message content (J. E. Arnold et al., 2004). Other aspects of manner of delivery such as tonal composition have also been shown to influence the pragmatic interpretation derived by listeners (Kurumada et al., 2014). In this thesis, we explore whether, and how, manner of delivery in the form of a disfluency in the utterance influences a listener's on-line pragmatic interpretation. We focus on two aspects of pragmatic interpretation: Scalar inferences, where meaning is ambiguous, and the perception of deception, where meaning may be the opposite of what is said. The former has been addressed from a time course perspective (Grodner et al., 2010; Huang & Snedeker, 2011; Tomlinson et al., 2013), but is under-explored with respect to a speaker's manner of delivery; the latter has been studied extensively in relation to manner of delivery (DePaulo et al., 2003; Zuckerman, Koestner, & Driver, 1981), but is under-researched in terms of the timing of the process. This thesis thus focuses on these two areas in attempt to address broader questions surrounding manner of delivery and the time course of pragmatic interpretation. Here, we provide a brief introduction on each of the two areas, focusing on how listeners derive the pragmatic meaning within each context.

2.5.1 Scalar inferences

Scalar inferences are a class of pragmatic inference associated with elements on a scale (Horn, 1984, 2004), where the use of a less informative element implies that a more informative term is not applicable. Take as a classic example the English quantifier *some*:

(9) a. Some coconuts grow on trees
b. Some, but not all, coconuts grow on trees
c. Some, and possibly all, coconuts grow on trees

Sentences such as (9a) typically evoke the interpretation (9b) (Gazdar, 1979; Horn, 1972; Van Tiel, Van Miltenburg, Zevakhina, & Geurts, 2016). Under a Gricean account, the inference from *some* to *not all* is said to occur because quantifiers such as *some* evoke a highly accessible set of alternatives that are ordered on a scale of informativeness (e.g., <some, all>). This inference reflects reasoning based on the maxim of Quantity (be informative): A cooperative speaker could have made (9a) more informative by using *all* instead; the fact that they didn't implies that they were not in a position to do so, triggering the pragmatic meaning of (9b). It should be noted that the Gricean theory of scalar inferencing has been challenged, notably by researchers who argue for a grammatical approach. Under this account, such inferences are stored in the lexicon and computed automatically by a dedicated grammatical system (e.g., Chierchia, 2004, 2006; Levinson, 2000; see Fox, 2014 for a review of the two approaches). Regardless of how the pragmatic meaning is derived, it has been argued that a core definition of *some* only limits its lower bound to *at least one* (and therefore can encompass all; Bott, Bailey, & Grodner, 2012; Breheny, Katsos, & Williams, 2006; Degen & Tanenhaus, 2014; Katsos, Breheny, & Williams, 2005). Hence, a strictly literal interpretation of (9a) is consistent with (9c). From a listener's perspective, this compatibility of *some* with both meanings generates ambiguity, whereby the quantifier can elicit either a literal or a pragmatic interpretation (e.g., Degen & Tanenhaus, 2014; Feeney, Scrafton, Duckworth, & Handley, 2004; Noveck, 2001). In this thesis, we address the resolution of this ambiguity in a context with extralinguistic cues.

2.5.2 Perceiving deception

Ambiguity in meaning may also arise when the intentions of a speaker are called into question. One such example is in the case of deception. Deception is broadly defined as a deliberate attempt to instil a false belief in a listener (Hala, Chandler, & Fritz, 1991). This can manifest in many forms, such as straightforward lying, double-bluffing, exaggerating, concealing the truth, among others (see Dynel, 2011 and Turner, Edgley, & Olmstead, 1975 for studies on various forms of deception). In this thesis, we focus on how listeners may perceive an utterance as a straightforward lie, where meaning is inferred to be the opposite of what is literally said (i.e. the true meaning of "it is X" is actually "it is not X"). This is the form of deception that lie perception studies typically assume listeners deduce (e.g., C. F. Bond, Omar, Mahmoud, & Bonser, 1990; C. L. Hart, Fillmore, & Griffith, 2009; Levine, Park, & McCornack, 1999; Vrij, Edward, & Bull, 2001). While an extended investigation of how interpretation may vary depending on what kind of deception is at stake lies beyond the scope of this thesis, we acknowledge that listeners may consider other forms of deception when told to expect a deceptive speaker (see, for example, sections 5.2.3 and 6.2.5), although ultimately our results do not support this hypothesis (see section 6.3).

Under a Gricean framework, deception is most aptly captured within the first of the four strategies by which a speaker can defy the Cooperative Principle, whereby they "quietly and unostentatiously violate a maxim" (Grice, 1975, p.49). For straightforward lies this depends on the maxim of Quality (be truthful), although different forms of deception can additionally violate different maxims (Gupta, Sakamoto, & Ortony, 2012). Deception therefore represents a deviation from the communicative standard set up by the Cooperative Principle, or to put it more bluntly, a form of pragmatic non-cooperation (Heyd, 2013).

For listeners, the first step to perceiving deception demands a recognition of the possibility of the speaker's non-cooperativity. This involves a process of pragmatic reasoning about the speaker's goals and intentions within the given situation (Franke, De Jager, & Van Rooij, 2009). There are many potential causes for a listener to infer deceptive intent in a speaker: Studies on the perception of deception suggest that listeners draw on various sources of information, from stereotypical behaviours such as disruptions in speech (Zuckerman, Koestner, & Driver, 1981) to local, context-specific cues like change in a speaker's behaviour (C. L. Hart et al., 2009; see DePaulo et al., 2003 for a review). These cues lead a listener to deduce that the speaker has violated a maxim, prompting them to adjust their view accordingly to derive a new set of beliefs. The outcome of this process is a rejection of the literal meaning, as the listener replaces it with an appropriate pragmatic interpretation (e.g., the opposite of what was said).

2.5.3 Investigating the time course of pragmatic inferencing

This thesis investigates listeners' pragmatic inferences during the comprehension of scalars and the perception of deception. For both contexts, we explore the time course of processing through a method that combines eye- and mouse-tracking within a visual world paradigm. These techniques have separately proven an effective and rewarding method to study the ongoing cognitive processes underlying comprehension (see Huettig et al., 2011 and Freeman, Dale, & Farmer, 2011 for reviews on eye-tracking and mouse-tracking respectively). The comprehension studies in this thesis employ the same basic design, in which listeners select between two objects on a display based on their interpretation of a speaker's utterance. One of the objects represents a literal meaning of the utterance, while the other evokes a pragmatic inference. We vary the speaker's manner of delivery by presenting fluent and disfluent forms of critical utterances. By comparing the pattern of eye and mouse movements during fluent and disfluent utterances, we gain an idea of how the speaker's manner of delivery guides listeners' pragmatic hypotheses in response to the unfolding speech.

Our results demonstrate that the manner of delivery has a rapid impact on listeners' pragmatic interpretation, in both the comprehension of scalars and the perception of deception (Chapters 3 and 4). In the case of deception, this effect is robust even in the face of other evidence about the speaker's dishonesty (Chapter 5 Experiment 5.1) or another potential cause of the speaker's disfluency (Chapter 5 Experiment 5.2). In the final study (Chapter 6), we investigate the production and perception of manner of delivery in a real-life, ecological context, by adapting the comprehension paradigm into an interactive game of deception. Here, we find a surprising result, where listeners still associate a speaker's filled pauses with deception, but speakers' pause behaviour suggests the opposite of what listeners seem to expect.

Chapter 3

Manner of delivery when meaning is ambiguous

Natural, everyday language is notoriously ambiguous. For example, words may be compatible with different meanings (e.g., "pen", "punch"; Duffy, Morris, & Rayner, 1988; Klepousniotou, 2002), sentences may have more than one syntactic parse (e.g., "The brother of the colonel who shot himself had been very depressed"; Traxler, Pickering, & Clifton, 1998), or expressions may be under-specific in what they refer to (e.g., "I've been *up north*", "Use *that thing* to remove the flywheel"; Appelt, 1985; Jucker, Smith, & Lüdge, 2003). Ambiguity also arises in the case of scalar expressions, which can elicit so-called literal or pragmatic interpretations. This is the form of ambiguity that this chapter is concerned with.

Studies on ambiguity resolution demonstrate that listeners may draw on various sources of information to converge on an interpretation (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey et al., 2002). One source that listeners can rely on is the speaker's manner of delivery (Bailey & Ferreira, 2007; Kjelgaard & Speer, 1999; Snedeker & Trueswell, 2003). For example, Bailey and Ferreira showed that disfluencies constrain the on-line parsing of syntactic ambiguity, by biasing listeners toward a modified noun phrase reading of a noun within a garden-path sentence (see section 2.4.1). In this chapter, we explore how spoken manner of delivery guides a listener's interpretation of ambiguity in a different context—disambiguation among possible meanings of the scalar quantifier *some*. In particular, we are interested in the relative time course with which the literal and pragmatic meanings develop in response to the unfolding speech.

3.1 The meaning of *some*

As noted in section 2.5.1, the scalar quantifier *some* can elicit two meanings: A pragmatic meaning referring to a subset of the total set of objects (*some but not all*), and a literal interpretation compatible with the totality of objects (*some and possibly all*). Ambiguity between the two alternatives has yielded a long line of research investigating how listeners derive the speaker's intended meaning (e.g., Breheny et al., 2006; Degen, 2015; Horn, 1984; Van Tiel et al., 2016), and the time course with which each meaning arises during comprehension (e.g., Bott & Noveck, 2004; Degen & Tanenhaus, 2014; Huang & Snedeker, 2011; Tomlinson et al., 2013; see Chemla & Singh, 2014a, 2014b for reviews). Although research shows that adult comprehenders typically favour an eventual pragmatic interpretation (Grice, 1975; Horn, 1972; Noveck, 2001; Papafragou & Musolino, 2003; Van Tiel et al., 2016), a continuing debate centres around when this occurs, and whether the literal meaning takes precedence, necessarily arising before the pragmatic inference.

A number of results within the literature suggest that the pragmatic interpretation of *some* is delayed relative to its literal counterpart, requiring more time and effort to derive (Bott & Noveck, 2004; Degen & Tanenhaus, 2011; Huang & Snedeker, 2011; Noveck & Posada, 2003; Storto & Tanenhaus, 2005). In a series of eye-tracking studies, Huang and Snedeker (2009a, 2009b, 2011) demonstrate that listeners consistently interpret *some* as compatible with a literal interpretation first before converging on the pragmatic interpretation. Similar processing costs have been observed in experiments using sentence verification tasks (Bott et al., 2012; Bott & Noveck, 2004) as well as self-paced reading (Breheny et al., 2006). Bott and Noveck showed, for example, that participants took longer to evaluate ambiguous sentences such as "Some animals are mammals" when instructed to assume a pragmatic interpretation of *some*, compared to those instructed to assume a literal interpretation. This delay associated with the pragmatic inference is consistent with predictions by the standard pragmatic model, which maintains a temporal distinction between slow pragmatic computations and rapid literal interpretations. This position, however, has not gone unchallenged. Using a similar paradigm and methods to Huang and Snedeker, Grodner et al. (2010) showed that the pragmatic meaning of *some but not all* can arise from the earliest stages of comprehension, with no evidence of precedence by literal comprehension (see section 2.3 for a detailed review of these studies). Grodner et al. argue that, with appropriate and sufficient contextual support, listeners can rapidly derive the pragmatic meaning of *some* without first having to compute a literal interpretation.

Grodner et al.'s results forefront the relevance of context in the comprehension of scalar expressions, a position similarly asserted by a number of other researchers (Bonnefon, Feeney, & Villejoubert, 2009; Breheny et al., 2006; Chierchia, 2004; Degen & Tanenhaus, 2014). Breheny et al. show, for instance, that the speed with which the pragmatic meaning is calculated is determined by the discourse context: In lower-bound contexts, where a reading of *at least one* is sufficient and an interpretation of *possibly all* holds no further informational value, participants' reading times were faster compared to upper-bound contexts, which encouraged a pragmatic interpretation of *some*. Although Breheny et al.'s results support a two-step approach to calculation of the pragmatic meaning, their results nevertheless make a case for the role of context in scalar comprehension. Bonnefon et al. (2009) demonstrate context-dependency using a different class of context: one in which the speaker's politeness goals may be relevant to a listener (Brown & Levinson, 1987). They showed that in cases where an utterance represents a face-threat to a listener (cf. Goffman, 1967), participants were less likely to endorse the pragmatic interpretation. For example, the sentence "Some people hated your poem" was more likely to give rise to the literal interpretation that *everyone* hated your poem. In contrast, a face-boosting version of the same utterance (e.g., "Some people loved your poem") was more likely to elicit the pragmatic interpretation that *not everyone* loved your poem. Bonnefon et al. suggest that in contexts where politeness concerns may be relevant to the discourse, individuals may construe the quantifier *some* as a politeness device employed by the speaker to mitigate the effects of a potential face-threat, leading listeners to consider the more face-threatening interpretation of the utterance. In the experiment reported in this chapter, we similarly target a social context based on the concept of face to investigate listeners' interpretation of *some*, going beyond the question asked in Bonnefon et al.'s study to explore the time course of this incorporation of context and its interaction with other cues from the speaker.

The above-mentioned studies provide evidence that scalar comprehension may depend on its global context of occurrence. However, the majority of studies to date have focused on straightforward comprehension of the scalar expression; comparatively little work has examined how this interpretation may be affected by a more local source of variation, that of the speaker's manner of delivery. Research outside of scalar comprehension, however, highlights the perceptual relevance of spoken manner of delivery on various aspects of pragmatic comprehension. Prosodic cues such as a speaker's pitch contour create expectations about an additional (unstated) meaning (Kurumada et al., 2014), while voice-based information evokes social inferences about a speaker's identity (Van Berkum et al., 2008). Disfluencies, such as filled pauses, also influence listeners' assessments of a speaker's veracity, with disfluent utterances biasing listeners toward inferring that a speaker is lying (King, Loy, & Corley, 2017; Loy, Rohde, & Corley, 2016a). Importantly, these studies provide evidence that listeners very quickly integrate these metalinguistic cues from the earliest moments of comprehension to shape their overall interpretation of the utterance. These findings highlight the speed with which manner of delivery can affect meaning construction via a process of social reasoning, undermining predictions by the standard pragmatic model that pragmatic enrichment only occurs at a relatively late stage of comprehension. It remains unknown, however, as to whether, and how, manner of delivery may influence a listener's real-time interpretation of an element that itself is associated with inferential processing: the ambiguous quantifier *some*.

Preliminary work by Bonnefon, Dahl, and Holtgraves (2015) suggests that manner of delivery may influence comprehenders' eventual interpretations of *some* within a relevant social context. Based on earlier findings that a speaker's politeness goals affect the meaning of *some* derived by comprehenders (Bonnefon et al., 2009), Bonnefon et al. (2015) hypothesised that silent pauses might influence listeners' interpretations of the quantifier within such a context, by functioning as a social cue to shift expectations toward unpleasant information. The study manipulated the description of whether or not a speaker remained silent in a scenario before delivering a face-threatening expression (e.g., 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"), and asked participants to rate on a scale of -5 to $+5$ the extent to which the statement warranted a literal interpretation (i.e. possibly everyone hated your idea). Their results showed that scenarios in which the speaker was described as remaining silent before speaking received higher ratings in favour of the more unpleasant interpretation—in this case the literal interpretation of *some*. Conversely, with a face-boosting expression (e.g., "Some people loved your idea") where the more unpleasant reading is the pragmatic meaning (i.e. not everyone loved your idea), the same pause description yielded higher ratings in favour of a pragmatic interpretation.

Bonnefon et al.'s results are indicative on two fronts. Firstly, and in line with previous work, they outline a relationship between context and the comprehension of scalar expressions (cf. Bonnefon, De Neys, & Feeney, 2011; Bonnefon et al., 2009; Breheny et al., 2006; Cummins & Rohde, 2015; Degen & Tanenhaus, 2014; Feeney & Bonnefon, 2012; Katsos & Bishop, 2011). Second, they provide prima facie evidence that given a relevant context, interpretations of *some* may vary with the manner in which the utterance is presented. However, their use of text-based stimuli, where the presence or absence of a pause was represented descriptively, limits the applicability of their results to real-life communicative situations which involve the auditory perception of utterances (and any accompanying hesitation). Furthermore, Bonnefon et al. used a task in which participants were explicitly asked to consider the possibility of the literal meaning of *some*. As such, their results likely reflect metalinguistic reasoning about a described event illustrating a speaker's delivery of the utterance. While these findings suggest a relationship between the manner of delivery of an expression and a comprehender's eventual considered interpretation of *some*, they leave open the question of whether such cues influence the interpretation of more naturally-produced utterances during real-time comprehension. In this chapter, we investigate whether spoken manner of delivery influences a listener's interpretation of the ambiguous quantifier *some*, during the moment-to-moment processing of the linguistic expression. Below, we present an experiment in which listeners make an implicit choice between a literal and a pragmatic interpretation of *some* based on a speaker's fluent or disfluent delivery of the scalar expression.

3.2 Experiment 3.1

The aim of the experiment was to investigate whether, and how, listeners' initial and final interpretations of *some* vary with the manner in which an utterance is conveyed. Building on previous work, we manipulated manner of delivery by using a disfluency in the form of a filled pause (e.g., Brennan & Williams, 1995; Corley et al., 2007). In a similar manner to Bonnefon et al. (2015), we established a context that exploited the concept of face (Goffman, 1967)—in this case, one in which a literal interpretation of *some* would threaten the positive self-image of the speaker. To achieve this, we invented a cover story about a fictitious experiment investigating greed and snacking habits.

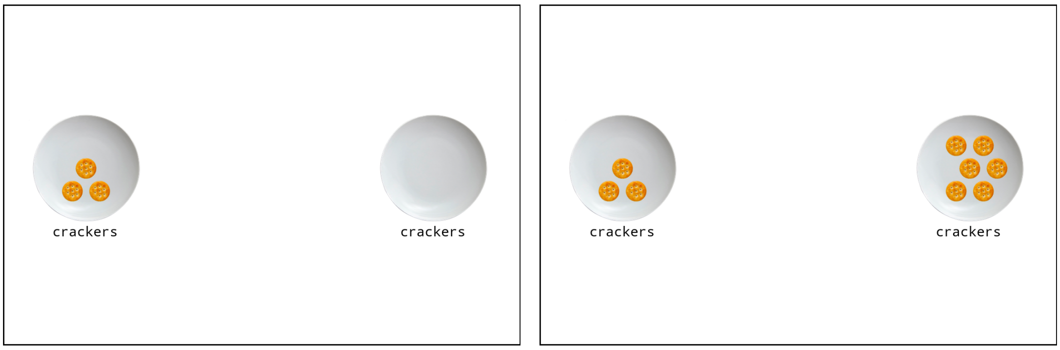
The cover story was as follows: We described a set of participants who were provided with a variety of snacks to eat while watching a documentary film. They received no instruction other than that they could eat as much or as little as they liked, and had to answer questions about the film in a verbal interview afterwards. We described that the study's motivation of investigating greed was subsequently revealed during the interview, and the fictitious participants were asked to report how much of each snack they had eaten (e.g., "I ate five crackers").

Participants in the current experiment were told that they would hear recordings of people who had taken part in the earlier experiment. This set up a context in which speakers who had consumed all the snacks might plausibly exploit the ambiguity of *some* to avoid face-loss through an admission to greed. Crucially, speakers might be disfluent as a by-product of the calculation of the potential threat to their positive self-image. In other words, "I ate *uh*, some oreos" could be taken to mean that a speaker ate *all* of the oreos but is embarrassed to admit it.

Each recorded utterance was played while participants viewed a visual display comprising two plates, with each plate depicting a quantity of one of the snack items. Participants were tasked with clicking on the plate that depicted what was left behind, based on the speaker's description. We measured their eye- and mouse-movements during each trial. Critical utterances made use of the quantifier *some*, with half of these including a filled pause disfluency.

We tested participants' interpretations of *some* using an Ambiguous display, where both plates were compatible with the utterance—one plate, with snacks remaining, corresponding to a pragmatic interpretation (*some but not all* eaten) and the other, empty, plate, to a literal interpretation (*all* eaten; see fig. 3.1).

Based on existing research, we expected an overall bias toward the pragmatic interpretation of *some* (Noveck, 2001; Papafragou & Musolino, 2003; Van Tiel et al., 2016). Importantly, on the basis that the face-saving context would induce listeners to interpret a disfluency as a signal that the speaker was avoiding face-loss, we expected filled pauses to yield a higher rate of the literal interpretation, and thus an increase in fixations on, mouse movements towards, and clicks on the empty plate.



(a) (b)

Figure 3.1: Example of displays used in (a) Ambiguous and (b) Control trials. In both displays the plate on the left depicts the pragmatic interpretation of *some*.

A potential concern with our predictions is that previous work has shown that listeners can interpret disfluency as a signal of simple deception, where the speaker means the opposite of what they say (Akehurst, Köhnken, Vrij, & Bull, 1996; Zuckerman, Koestner, & Driver, 1981). These effects have been shown to rapidly influence comprehension, emerging almost as soon as a listener can infer meaning based on the unfolding linguistic input (Loy et al., 2016a; see also Chapters 4 and 5). Thus, given the typical inference that *some* means *not all*, a disfluency construed as a signal of deception might quickly bias listeners toward the plate associated with the atypical literal meaning, in this case the empty plate. To rule out deception as a potential account of our findings, we included a Control display where only one plate, with a few snacks remaining, was compatible with any meaning of *some*. In this condition, the second plate—a distractor—always had 6 snacks remaining, corresponding to only one piece having been eaten (see fig. 3.1). Under a disfluency-signals-simple-deception hypothesis, disfluent utterances would bias listeners towards the distractor plate, because it is not compatible with the interpretation that the speaker has eaten *some* of the snacks. On the other hand, under the assumption that given a relevant social context, disfluency is associated with face-saving rather than deception, both fluent and disfluent utterances would result in fixations and clicks on the sole plate that is compatible with any meaning of *some*. Thus an absence of disfluency effect in the Control display would allow us to conclude that any effect of disfluency in the Ambiguous display is driven by listeners' context-dependent interpretation of *some*, rather than their establishment of an unstated meaning that is incompatible with any acceptable meaning of the quantifier.

Participants' eye- and mouse-movements were recorded on each trial, as well as their eventual interpretations (plate clicked) and response times. Central to our interest is the variation in bias toward the pragmatic interpretation depending on the manner of the utterance. If disfluency encourages a literal interpretation in face-saving contexts such as this, we would expect the literal plate on Ambiguous displays to function as a direct competitor. Hence, we would expect a lower proportion of pragmatic responses for Ambiguous compared to Control displays, where the competitor is not compatible with any interpretation of *some*. If this inference occurs during real-time comprehension, we would expect to see a reduction in the fixation and mouse-movement bias toward the pragmatic plate following disfluent utterances. For Control displays, disfluency should have little effect on listeners' responses, unless disfluency serves to signal simple deception rather than any plausible speaker embarrassment over the threat to their self-image set up by the face-saving context, in which case a bias away from the pragmatic plate is predicted.

3.2.1 Method

Participants

Twenty-four self-reported native British English speakers took part in the experiment. All were right-handed mouse users with normal or corrected-to-normal vision. An additional 3 participants were tested, but their data were not included in our analyses because they suspected the authenticity of the cover story (2) or that the audio had been scripted for the experiment (1; determined during debrief). Participants were recruited from the University of Edinburgh community and each received £5 for participation. All participants provided informed consent in accordance with the university's Psychology Research Ethics Committee guidelines (ref no.: 136-1617/1).

Materials and design

Eight different types of snacks were used as referents in the experiment. The cover story established a starting quantity of 7 for each snack item (see fig. 3.2). The 8 snacks were chosen based on a pre-test of 12 snacks, in which respondents indicated the likelihood that they would eat up to 7 pieces of each snack in one sitting.

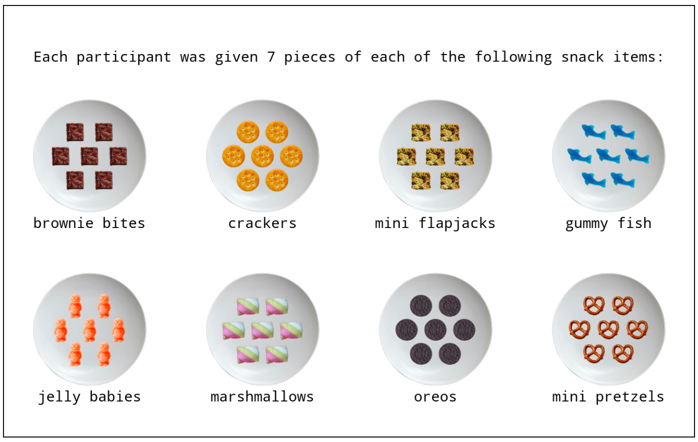


Figure 3.2: Snack items used as referents in the experiment

On each trial, participants saw a visual display comprising two plates, each depicting a quantity (range: 0–7) of one of the snack items. This quantity represented the number of pieces of the snack that remained (out of 7). The name of the snack was displayed below each plate to avoid ambiguity in cases where 0 pieces remained. Each display was accompanied by a recording of a speaker describing how much of a snack they had eaten. The utterances were produced by 8 speakers (4 male; all native British speakers), each contributing 8 utterances (one per snack), for a total of 64 utterances used in the experiment. Two out of each speaker's 8 utterances were critical utterances; the other 6 were fillers. Snacks were balanced across speakers such that each snack only occurred as a referent in two critical utterances, each by a different speaker, with no two speakers associated with the same two critical referents.

On critical utterances, the speaker used *some* to describe how much of the snack they had eaten. In Ambiguous displays, this was compatible with both plates—a pragmatic interpretation depicting 2, 3 or 4 remaining pieces of the referent (i.e. corresponding to 5, 4 or 3 pieces having been eaten), and a literal interpretation depicting 0 pieces. These quantities for the pragmatic interpretation were chosen based on evidence that *some* is perceived as most natural when used to reference intermediate set sizes (e.g., 6–8 out of 13 gumballs; Degen & Tanenhaus, 2014). In Control displays, *some* was compatible with only one of the two display plates—a pragmatic interpretation depicting 2, 3 or 4 remaining pieces. The second plate, a distractor, contained 6 pieces of the referent to illustrate one piece having been eaten—an interpretation intended to be incompatible with any meaning of *some*. Half of the utterances accompanying each display were fluent ("I ate some crackers") and the other half disfluent ("I ate uh, some crackers"). Hence, the study followed a 2 (manner: fluent/disfluent) x 2 (display: Ambiguous/Control) within-subjects design, with critical utterances counterbalanced across 4 lists. The quantity displayed on the pragmatic plate on each trial was chosen at random from a list, with 2, 3 and 4 represented equally across conditions. Within each condition, the pragmatic plate appeared on the left and the right an equal number of times.

The 16 critical trials were randomly presented together with 48 fillers. To increase variability, these included a number of manipulations in the speaker's manner. Half of the filler utterances were fluent; the other half contained some form of disfluency or a hedge suggesting uncertainty about the exact quantity eaten (see table 3.1). Filler utterance types were distributed across speakers such that each speaker produced an even mix of default and other utterances. Filler trials also varied at the level of display. In half the filler displays, the distractor plate depicted a different quantity of the referent snack. In the other half, the distractor plate depicted a different snack, with half of these depicting the same quantities of each snack. This manipulation had the purpose of discouraging listeners from focusing only on the determiner to disambiguate between the two plates on each trial. Filler displays were distributed such that each of the 8 speaker's filler utterances were accompanied by a variety of filler display types. The same set of filler trials was used in all 4 experimental lists.

Table 3.1: Breakdown and examples of filler utterances

Filler type	Example	No. of utterances
Fluent	I ate three crackers.	24
Disfluent (prolongation)	I *aate...* four oreos.	4
Disfluent (correction)	I ate *four- no,* five marshmallows.	4
Hedge	I *think* I ate six brownies.	4
Hedge	I ate *maybe...* three jelly babies.	4
Hedge	I ate *like,* two gummy fish.	4
Hedge	I ate *one or two* flapjacks.	4

Procedure

The experiment was presented using OpenSesame 3.1.0 (Mathôt, Schreij, & Theeuwes, 2012) on a 21in. CRT monitor. Eye movements were monitored using an Eyelink 1000 Tower Mount system sampling at 500Hz. Mouse coordinates were sampled at 50Hz.

Participants were first briefed on the cover story which established the context in which the utterances were produced. To corroborate the story, the instructions included a photo ostensibly taken of a participant taking part in the fictitious experiment.

Following the instructions, the eyetracker was calibrated. Between trials, participants underwent a manual drift correction using a central grey fixation dot. After this, the dot turned red for 500 ms to signal the start of the trial. Each trial began with a 1000 ms presentation of the two full plates containing 7 items each. This served to remind participants of the starting quantity of each snack item. The two plates were centred vertically and positioned horizontally left and right on the screen. This was followed by a 1000 ms preview of the actual quantities associated with each snack for the trial. After this, a mouse pointer appeared at the centre of the screen and playback of the utterance began. Participants were instructed to click on the plate depicting the quantity remaining based on the speaker's description of what they ate. For example, if the participant heard "I ate five oreos", they would click on the plate depicting two oreos. There was no feedback except in cases where participants failed to click on a plate within 5000 ms post-utterance offset, following which they received a message to respond more quickly. Participants underwent 4 practice trials and were given the opportunity to ask questions afterwards, before the main experiment began. None of the practice trials contained *some* in the utterance.

3.2.2 Results

Statistical analyses were carried out in R Version 3.3.3 (R Core Team, 2017). Our analyses focused on listeners' final interpretations of *some* for each utterance (plate clicked), response times, eye movements and mouse movements. For each dependent variable, we modelled the effect of manner of delivery (fluent/disfluent) individually for each display (Ambiguous/Control). To evaluate the difference in the effect of manner on the two displays, we also ran an interaction model taking into account both manner and display as fixed effects. Predictors were mean centred in all analyses.

Logistic regression was used to model the binary outcome of which plate participants clicked on. The distribution of responses reflected an overwhelming bias toward a pragmatic interpretation of *some*. To avoid spurious ceiling effects, a generalised linear model by robust methods was fit using the glmrob function from the robustbase package (Maechler et al., 2016). This approach produces more robust estimation of regression parameters in cases where inference based on maximum likelihood may yield unreliable results (Cantoni & Ronchetti, 2001). Linear mixed effects regression was used to model participants' response times, using the lmer function from the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). Models included by-subjects and by-items random intercepts and slopes for manner and display.

Eye-tracking records were averaged into 20 ms bins, each comprising 10 samples, prior to analysis. Data were coded in terms of fixations toward either one of the plates or areas outside of both. The proportion of fixations to each plate out of the total sum of fixations was computed for each time bin. Mouse-tracking analysis only took into account the X coordinates. For each sample, the distance travelled by the mouse was computed by taking the absolute difference between the X coordinates of the current and previous samples. The data were coded for direction of movement toward either one of the plates for each bin, and the cumulative distance participants had moved the mouse toward each plate was computed by summing over the distance travelled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). For each plate, we then calculated a proportion-of-movement measure, defined as the distance travelled by the mouse pointer towards the given object, divided by the total distance travelled (regardless of X direction).

To evaluate whether manner of delivery influences listeners' processing of *some* during real-time comprehension, eye- and mouse-tracking data were analysed over an 800 ms time window beginning from 200 ms post-quantifier onset. This window corresponds to the duration of the quantifier and subsequent referent, taking into account the 200 ms it typically takes to program and execute an eye movement (Matin, Shao, & Boff, 1993), and ending just before the average utterance offset (1071 ms). Models for this window were fitted using empirical logit regression (Barr, 2008), taking as the dependent variable the difference between the e-logit of fixations (or mouse movements) to the two plates on each trial. Fixed effects included time, manner and display (all predictors mean centred). All models included by-subjects and by-items random intercepts and slopes for all predictors.

Click responses

Table 3.2 shows the breakdown of mouse clicks recorded on each plate following fluent and disfluent utterances on each display. The last column shows the mean response time (in ms) measured from the onset of *some*.

For Ambiguous displays, a robust logistic regression on the outcome of mouse clicks showed an effect of manner of delivery. Disfluent utterances resulted in fewer clicks on the pragmatic plate (and therefore more clicks on the literal plate), $\beta = -1.70$, SE $= 0.86$, $p = .05$. A linear mixed effects regression on listeners' response times showed an effect of manner of delivery. Listeners were slower to click on a plate following a disfluent utterance, $\beta = 260.46$, SE $= 74.81$, $t = 3.48$. For Control displays, there was no effect of manner on listeners' mouse clicks

Table 3.2: Breakdown of mouse clicks (raw count) recorded on each plate and mean response times (in ms) following fluent/disfluent utterances on Ambiguous/Control displays. Standard errors are in parentheses.

		plate clicked		response time
		pragmatic	competitor[a]	
Ambiguous	Fluent	94	2	1842.40 (74.56)
	Disfluent	86	10	2102.86 (72.98)
Control	Fluent	95	1	1914.29 (66.03)
	Disfluent	94	2	1806.60 (63.62)

[a] On Ambiguous displays the competitor plate represented the literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*.

(p = .6) or response times ($t = -1.38$). These results provide no evidence to support the disfluency-signals-simple-deception hypothesis.

A robust logistic regression on listeners' mouse clicks including both manner and display as predictors showed no effect of either, nor any interaction between the two (all $p > .1$). A linear mixed effects regression on response times showed no main effects of manner or display, but yielded a manner by display interaction, $\beta = 368.15$, SE = 106.78, $t = 3.45$, reflecting the longer time taken by listeners to click on a plate following a disfluent utterance on Ambiguous displays.

Eye movements

Fig. 3.3 shows the proportion of fixations to each plate over time until 2,000 ms post-*some* onset, by which point participants had typically moved the mouse over one of the two plates. The pattern of fixations on Ambiguous displays demonstrates a baseline bias toward the pragmatic plate relative to the literal plate. This likely reflects a preference to look at plates with objects over empty plates,

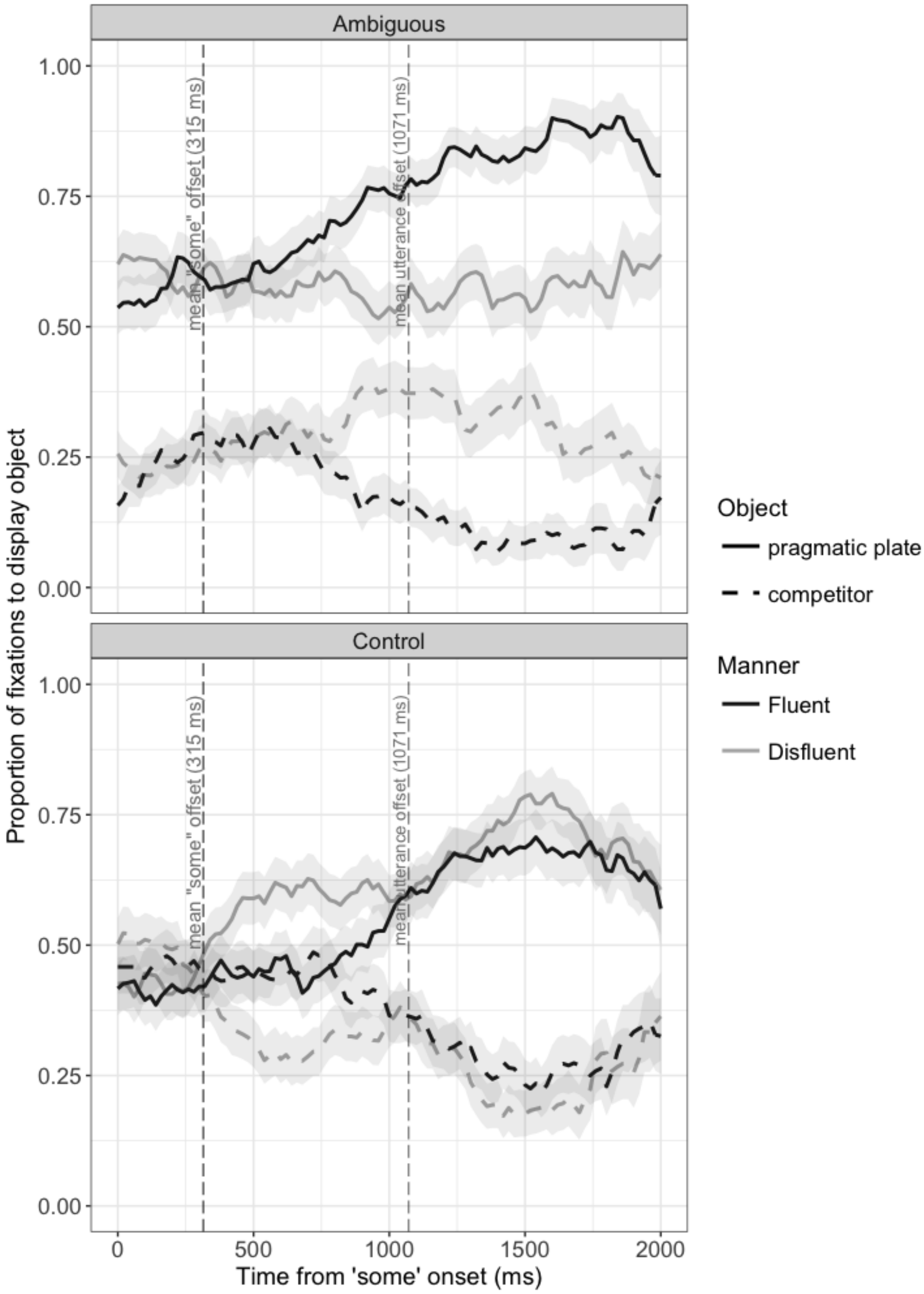


Figure 3.3: Proportion of fixations to each display plate over time during fluent and disfluent utterances for Ambiguous (top) and Control (bottom) displays. Shaded areas represent ± 1 SE of the mean. On Ambiguous displays the competitor represented a literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*.

and is consistent with earlier studies which report a fixation bias to the image with the largest quantity of items prior to disambiguation (Grodner et al., 2010; Huang & Snedeker, 2009b). As predicted, under a model in which disfluency is interpreted in a social context, the time course of fixations on Ambiguous displays suggests a difference dependent on the manner of delivery. Fluent utterances led to a rapid rise in fixations to the pragmatic plate after the onset of *some*; on disfluent utterances, this increase was attenuated. This difference was reflected in a time by manner interaction, $\beta = 8.11$, SE = 1.05, $t = 7.76$.

In contrast, on Control displays, disfluent utterances saw an earlier rise in fixations to the pragmatic plate compared to fluent utterances, as evidenced by a time by manner interaction, $\beta = -2.67$, SE = 1.13, $t = -2.27$. The difference in the effect of manner on the Ambiguous and Control displays was confirmed by a three-way time by manner by display interaction, $\beta = 10.88$, SE = 1.53, $t = 7.11$. Although the effect of manner on Control displays was unexpected, we note that this result nevertheless does not support the disfluency-signals-simple-deception hypothesis, which predicts a fixation bias to the competitor plate following disfluent utterances. We return to this effect in the Discussion.

Mouse movements

Fig. 3.4 shows the proportion of mouse movements (in terms of distance travelled) toward each plate over time until 2,000 ms post-*some* onset. The time course of mouse movements follows a pattern compatible with the fixation data. On Ambiguous displays, participants' mouse movements exhibited a greater preference for the pragmatic plate over the literal plate following fluent utterances, while this increase was attenuated during disfluent utterances, $\beta = 5.03$, SE = 0.61, $t = 8.25$. In contrast, on Control displays mouse movements were characterised by a greater preference for the pragmatic plate over the competitor during disfluent utterances, $\beta = -1.70$, SE = 0.57, $t = -2.96$. This effect aligns with the early fixation bias to the pragmatic plate following disfluent utterances on

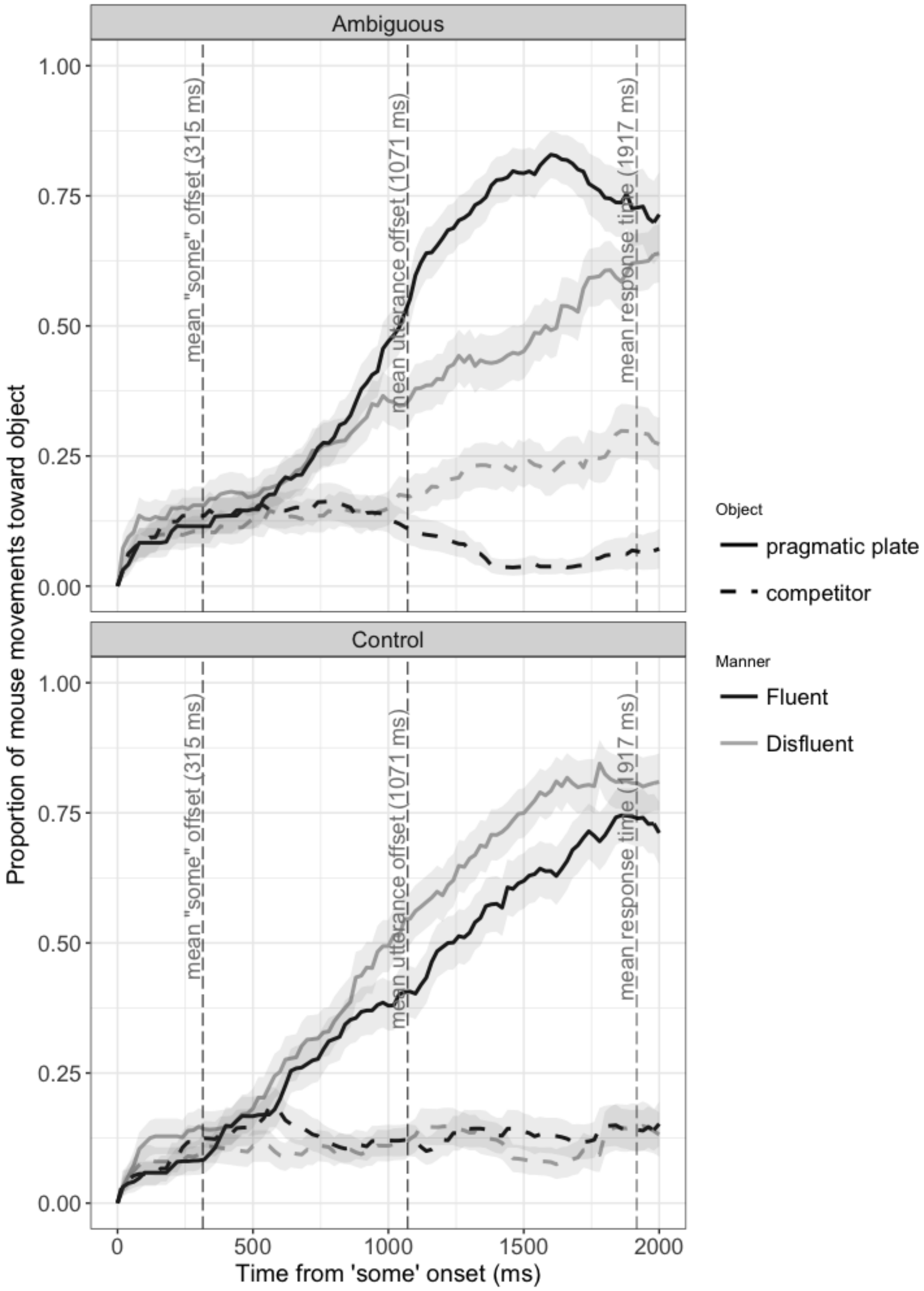


Figure 3.4: Proportion of mouse movements to each display plate over time during fluent and disfluent utterances for Ambiguous (top) and Control (bottom) displays. Shaded areas represent ±1 SE of the mean. On Ambiguous displays the competitor represented a literal meaning of *some*; on Control displays the competitor was a distractor incompatible with any meaning of *some*.

Control displays. As with the eye movements, the difference in the effect of manner on listeners' mouse movements during the two displays was confirmed by a three-way time by manner by display interaction, $\beta = 7.22$, SE = 0.87, $t = 8.34$.

3.2.3 Discussion

In this chapter, we set out to test whether listeners' interpretations of the ambiguous quantifier *some* varies with the speaker's manner of delivery. Like Bonnefon et al. (2015), we made use of a social context that exploited the concept of face—in this case one where snacking is associated with greed, which in turn threatens the positive self-image of a speaker. This allowed us to establish a context in which a speaker's disfluency could be perceived as a social cue that signals this potential face-loss for the speaker. Our results suggest that listeners did indeed assign this social meaning to speakers' disfluencies. Fluent utterances yielded an overwhelming bias toward the pragmatic interpretation. This pattern follows a robust trend established in the literature for adult listeners to assign to *some* a meaning of *not all*. However, when the literal meaning—the plate associated with the socially dispreferred meaning of having greedily eaten all the snacks—was available as an alternative interpretation, disfluency reduced the bias toward the pragmatic interpretation. This was apparent in Ambiguous displays, where disfluent utterances led to a decrease in the proportion mouse clicks on the pragmatic plate in favour of the literal plate.

Moreover, the time course of eye and mouse movements show that this effect of manner emerged during the early stages of comprehension, almost as soon as listeners could establish a meaning of *some*, and prior to completion of the speaker's utterance. This finding supports a growing literature demonstrating how manner of delivery can modulate various aspects of a listener's on-line pragmatic hypotheses (e.g., Kurumada et al., 2014; Van Berkum et al., 2008). Here, we show how the presence of disfluency in an utterance reduces listeners' tendency to derive a pragmatic meaning of *some*. Moreover, this effect was not the consequence of perceiving the disfluency as a context-independent signal of deception. This is evident from results from the Control condition. Here, under a simple deception account for the effect of manner, disfluent utterances should have elicited a bias toward the competitor—the plate depicting a quantity incompatible the speaker's utterance. However, we found no evidence of such a bias in Control displays. Listeners' mouse clicks on the pragmatic plate were at ceiling for both fluent and disfluent utterances, while their eye and mouse movements suggest that disfluent utterances in fact led to an earlier bias toward the pragmatic plate compared to fluent utterances. These results allow us to rule out a simple deception account for the effect of manner on the interpretation of the scalar.

Instead, listeners' behaviours across the two displays suggest that they very quickly interpreted speakers' disfluencies within the global discourse context to shape their overall interpretation of the utterance (cf. Bonnefon et al., 2015). That is, given the social context of the fictitious experiment whereby a speaker may be motivated to protect their positive self-image, disfluency likely signalled the speaker's discomfort due to a potential threat to this self-image, prompting listeners to favour a more self-image-threatening interpretation of the utterance. Accordingly, when afforded both the literal and pragmatic interpretations of *some* in Ambiguous displays, disfluent utterances encouraged listeners to consider the literal meaning (where *in fact all* of the snacks had been eaten), resulting in the decreased commitment to the pragmatic meaning relative to fluent utterances. Notably, such a context-based account of the effect of disfluency would also explain the earlier bias toward pragmatic-*some* following disfluent utterances observed in Control displays—expectation of a potential face-loss for the speaker could have led listeners to infer that a larger quantity of snacks had been eaten, resulting in an anticipatory bias toward the pragmatic plate over the competitor, in which only one piece had been eaten. Within the field of scalar research, our findings are consistent with the view that the interpretation of *some* depends on the context in which it occurs (Bonnefon et al., 2015, 2009; Breheny et al., 2006;

Cummins & Rohde, 2015; Degen & Tanenhaus, 2014; Grodner et al., 2010). The majority of these studies have focused on the role of context in the off-line, eventual interpretation of scalar expressions. However, in light of our current results, we suggest that future research should take into account contextual-dependency in examining the time course of scalar comprehension.

The results of this experiment allow us to advance two main findings with respect to the scope of the thesis. First, extending findings from the disfluency literature that listeners make rapid use of a speaker's manner of delivery to evaluate syntactic ambiguity (Bailey & Ferreira, 2007) or to predict semantic content (J. E. Arnold et al., 2007, 2004; Barr & Seyfeddinipur, 2010), we demonstrate disfluency's role as a social cue in guiding listeners' on-line pragmatic inferences. Taken together, these results highlight the flexibility of the comprehension system in using manner of delivery as a cue to facilitate understanding, by drawing on different processes depending on the comprehension goals of the listener. More broadly, we show that listeners' pragmatic hypotheses about a speaker's utterance unfold during the initial stages of comprehension. In particular, we found no evidence to support a temporal precedence of literal comprehension before the pragmatic interpretation. This latter finding runs counter to predictions by the standard pragmatic model, which maintains that listeners must first analyse the literal meaning of utterances before any form of pragmatic enrichment can take place. Contrary to this, our results indicate that listeners' pragmatic inferences about the meaning of *some* unfolded immediately, and were rapidly modulated by a cue in the speaker's manner of delivery within the given social context.

In sum, our results from this experiment reveal that manner of delivery, in the form of utterance fluency, can very quickly constrain a listener's pragmatic inferences in the context of assigning meaning to the ambiguous quantifier *some*. This raises questions about the relevance of this manner-based cue in other pragmatic contexts, where listeners may have to similarly reason about non-literal meanings to derive a contextually-appropriate interpretation. In the next chapter, we explore this question by turning our attention to a different aspect of pragmatic comprehension, in a context where listeners have to make an inference about the veracity of an utterance based on their perception of the speaker's intentions.

Chapter 4

Manner of delivery when a speaker is unreliable[1]

In the previous chapter, we established that listeners rapidly formulate hypotheses about the meaning of an ambiguous quantifier *some* based on the speaker's manner of delivery. Specifically, in a context where the positive face of the speaker is at stake, fluent utterances quickly bias listeners toward a pragmatic interpretation of *some*, while disfluency attenuates this bias in favour of the literal interpretation. Importantly, this bias emerges during the initial stages of comprehension, within hundreds of milliseconds from the point at which listeners can assign a meaning to *some*. Beyond demonstrating the role of manner of delivery in the comprehension of scalars, this finding highlights that pragmatic understanding is fundamentally a contextually rich and rapid process, with linguistic meaning highly coloured by pragmatics from the earliest stages of comprehension. In the

[1]This chapter constitutes an extended version of a published paper (Loy, J. E., Rohde, H. & Corley, M. (2016) Effects of disfluency in online interpretation of deception. *Cognitive Science.* doi: 10.1111/cogs.12378). In the current chapter, figures were recompiled to maintain consistency with the formatting used throughout this thesis. There are therefore some superficial differences between the versions presented here and in the original paper. In all other respects the content of the paper remains unchanged.

empirical chapters that follow (4–6), we turn to pragmatic comprehension in a different context, one in which the meaning of an utterance may change based on the perceived intentions of the speaker. We focus on the perception of deception, where listeners derive an alternative, non-literal meaning of an utterance via an inference that the speaker is lying. We investigate how manner of delivery functions as a pragmatic cue to deception (Chapter 4), additional factors which may influence the perceptual relevance of this cue (Chapter 5), as well as whether listeners' expectations based on the speaker's manner of delivery align with speakers' actual production of cues to deception (Chapter 6). In this chapter, we first provide a broad overview of the perception of deception, focusing on the role played by a speaker's manner of delivery. We then present two experiments that examine whether listeners' pragmatic inferences about utterance veracity are influenced by its manner of delivery.

4.1 The perception of deception

Previous research on deception has typically addressed questions such as how accurately people are able to detect lies (e.g., C. L. Hart et al., 2009; Mann, Vrij, & Bull, 2004; see C. F. Bond & DePaulo, 2006 for a review), or the specific cues that are used to recognise a lie (e.g., Akehurst et al., 1996; Granhag, Strömwall, & Hartwig, 2005; Vrij, Edward, Roberts, & Bull, 2000; see Zuckerman, Koestner, & Driver, 1981 for a review). Despite the frequent regularity with which lying occurs in everyday discourse (DePaulo & Kashy, 1998; DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), studies demonstrate that people display unimpressive success at detecting falsehood. In a meta-analysis of 206 studies, C. F. Bond and DePaulo (2006) report a 54% truth-lie discrimination accuracy, or just slightly above what we'd expect based on chance alone. Researchers also commonly observe that people achieve greater success at correctly classifying truths as truthful than lies as deceptive (C. F. Bond & DePaulo, 2006; Levine et al., 1999; Vrij, 2000).

This finding has been attributed to a pre-existing *truth bias* (Vrij & Baxter, 1999), or the inclination for listeners to perceive utterances as truthful rather than deceptive. This may reveal a degree of expectation that speakers adhere to the Cooperative Principle (Grice, 1975), although it has been noted that it could reflect a more straightforward *availability heuristic*, by which people assume the truth based on greater exposure to truthful utterances in daily life (O'Sullivan, Ekman, & Friesen, 1988).

In terms of cue perception, research is largely founded on the premise that various psychological processes underlie the act of constructing a lie. These processes shape the emotional or mental state of a liar, which in turn affects certain aspects of their manner of delivery. These are the cues which listeners are assumed to draw on when inferring lie behaviour (Zuckerman, DePaulo, & Rosenthal, 1981). Three broad perspectives are frequently cited to explain how deception impacts a speaker's manner of delivery. The *emotional hypothesis* focuses on various emotions that a liar experiences, namely guilt, fear or excitement (Ekman, 1992). This may result in cues indicative of these emotions; for example, lying is frequently associated with nervous gestures such as gaze aversion or a higher pitched voice (Akehurst et al., 1996; C. F. Bond et al., 1990; Sporer & Schwandt, 2006). The *cognitive hypothesis* emphasises that deception is a cognitively complex task, resulting in cues that reflect this additional demand (Vrij, 2000; Vrij & Heaven, 1999; Vrij et al., 2008). Consequently, utterances containing speech errors or disfluencies are more likely to be interpreted as lies (DePaulo, Rosenthal, Rosenkrantz, & Green, 1982; Zuckerman, Koestner, & Driver, 1981), presumably due to the increased mental load of formulating a lie (Gombos, 2006). The *attempted control hypothesis* focuses on the impression management measures that liars may employ to maintain a credible appearance (Buller & Burgoon, 1996; DePaulo, 1988). This hypothesis suggests that liars attempt to suppress potential signs of deception by controlling their behaviour, but may in the process come across as unnatural. In line with this, some researchers have noted that liars often display an unusual degree of rigidity and inhibition (e.g., Vrij, 1995).

While research has identified a range of potential cues to deception, disagreement arises as to which cues listeners reliably use to discriminate truth from lie, as well as the direction of these relationships. For example, while some researchers report that an increase in voice pitch is used to index lie behaviour (Sporer & Schwandt, 2006; Zuckerman, Koestner, & Driver, 1981), others fail to find any evidence for this (Vrij & Semin, 1996). This inconsistency may in part be due to the spectrum of lies that listeners are asked to judge across lie perception studies (e.g., total falsehoods, exaggerations, withholding of information, etc.; see DePaulo et al., 1996; Dynel, 2011 for examples of classification systems). Different forms of lie may trigger distinct reasoning processes (cf. Gupta et al., 2012), increasing or decreasing the salience of certain manner-based cues. C. L. Hart et al. (2009) additionally propose that cue production may be idiosyncratic and specific to a liar. They show that listeners who indirectly detect liars by looking for changes in their speech or mannerisms achieve higher accuracy at classifying liars and truth-tellers than listeners explicitly asked to detect liars. C. L. Hart et al. suggest that it may be more relevant for listeners to focus on changes in manner of delivery within a given speaker, rather than to rely on a particular behaviour or set of behaviours. These findings highlight the challenges that researchers may face in establishing cue behaviours that generalise across situations.

Although researchers remain divided on the discriminative value of many cues, certain statistically reliable correlates have emerged across studies. Meta-analytic studies highlight that speakers' disfluencies, in particular their filled pauses, are frequently used by listeners to infer lie behaviour (Sporer & Schwandt, 2006; Zuckerman, DePaulo, & Rosenthal, 1981). This pattern corroborates evidence from questionnaire-based studies investigating beliefs about cues to deception, which suggest that filled pauses are among a set of cues that listeners associate with deception, whether in predicting behaviour exhibited by themselves or by others when lying (Akehurst et al., 1996; Zuckerman, Koestner, & Driver, 1981).

Studies also suggest that listeners tend to rely on certain nonverbal aspects of a speaker's manner to infer deception. In particular, increased fidgeting, such as hand, foot, and leg movements, emerge as reliable indicators of lie behaviour across studies (Hartwig & Bond, 2011; Zuckerman, Koestner, & Driver, 1981). This association is consistent with the assumption under a cognitive hypothesis that liars experience difficulty (Vrij et al., 2008), which may in turn manifest in the form of nonverbal behaviour in response to an increased mental load (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001).

It is worth noting that these associations still do not hold universally across studies. For example, C. F. Bond et al. (1990) observed that listeners' deception judgements did not vary with a speaker's filled pauses. They were, however, influenced by another form of disfluency—speakers' silent pauses, with an increase in this type of pause associated with lie behaviour. Nevertheless, while researchers remain uncertain over which cues hold perceptual relevance, the overall picture is clear: Listeners make use of cues in a speaker's manner of delivery to infer whether a speaker is lying. These associations suggest that listeners engage in a form of speaker modelling via the cognitive hypothesis, where lying is cognitively difficult and cognitive difficulty induces a change in the speaker's manner, perhaps in the form of a disfluency or a display in their body language. In this way, the speaker's manner of delivery provides useful information that listeners exploit via a process of reasoning to form pragmatic inferences about an utterance's veracity.

Although extensive research has documented this relationship between manner of delivery and lie perception, less is known about the *process* by which these cues inform listeners' inferences. The majority of lie perception studies have relied on off-line measures of judgement, such as surveying people about their explicit beliefs regarding cues to deception (e.g., Akehurst et al., 1996; Granhag et al., 2005), or asking participants to classify utterances as truths or lies (e.g., Vrij, 1993). These methods delineate an implicit assumption that pragmatic inferences, such as reasoning about a speaker's deceptive intent, occur relatively late in processing. This is a view consistent with the one held by the standard pragmatic model. However, such methods ultimately reveal little about listeners' unfolding inferential processes, or how they may be influenced by various external factors.

The following studies in this thesis represent a step toward addressing these questions. By extending our eye- and mouse-tracking paradigm to examine contexts in which listeners entertain the possibility of deception, we explore the time course with which their inferences about utterance veracity unfold, and how these are guided by the speaker's manner as well as other factors. In the remainder of this chapter, we present two experiments investigating the effect of manner of delivery on listeners' pragmatic inferences about utterance veracity. We modify the paradigm from Chapter 3 to elicit an implicit judgement about a speaker's truthfulness based on the fluency of the speaker's delivery. We also explore whether the location of disfluency matters to listeners (cf. Clark & Fox Tree, 2002), by using disfluencies at the start of the utterance (Experiment 4.1) as well as in the middle (Experiment 4.2).

4.2 Experiments 4.1 and 4.2

4.2.1 Abstract

A speaker's manner of delivery of an utterance can affect a listener's pragmatic interpretation of the message. Disfluencies (such as filled pauses) influence a listener's off-line assessment of whether the speaker is truthful or deceptive. Do listeners also form this assessment during the moment-by-moment processing of the linguistic message? Here we present two experiments that examined listeners' judgment of whether a speaker was indicating the true location of the prize in a game during fluent and disfluent utterances. Participants' eye and mouse movements were biased toward the location named by the speaker during fluent

utterances, whereas the opposite bias was observed during disfluent utterances. This difference emerged rapidly after the onset of the critical noun. Participants were similarly sensitive to disfluencies at the start of the utterance (Experiment 1) and in the middle (Experiment 2). Our findings support recent research showing that listeners integrate pragmatic information alongside semantic content during the earliest moments of language processing. Unlike prior work which has focused on pragmatic effects in the interpretation of the literal message, here we highlight disfluency's role in guiding a listener to an alternative non-literal message.

4.2.2 Introduction

In interpreting an utterance, listeners must often take into account not only *what* is said, but *how* it is being said, because the manner of production may in turn be relevant to the interpretation of meaning. For instance, a speaker may place a prosodic contour on a statement to emphasize a point (Fernald & Mazzie, 1991), or they may preface their utterance with an "um" to express a lack of certainty (Smith & Clark, 1993). Such paralinguistic features encoded within the speech signal are known to influence listeners' pragmatic interpretations of the linguistic message. In their study on the perception of speaker confidence, Brennan and Williams (1995) observed that utterances produced with rising intonation, temporal delays, and filled pauses such as "uh" or "um" were judged to be less confident, showing that the manner of delivery of an utterance provides cues that influence a listener's estimation of the speaker's metacognitive state. Brennan and Williams's findings have since been extended to a variety of contexts, such as the assessment of a speaker's certainty (Swerts & Krahmer, 2005), by both adult and child judges (Krahmer & Swerts, 2005), as well as in machine interpretation of nonspontaneous speech (Pon-Barry & Shieber, 2011).

Research on the recognition of deception also highlights the perceptual relevance of paralinguistic information in language comprehension. Studies on lie detection focus on the interpretation of paralinguistic cues in both auditory and visual modalities. Vocal cues, such as tone of voice or speech disturbances, appear to be more informative than visual cues such as facial gestures or body movements. Across an analysis of 50 studies that manipulated participants' access to visual and audio information when making truth-lie judgments, C. F. Bond and DePaulo (2006) found that discrimination accuracy was lower when participants made judgments based on visual information rather than audio or audiovisual, highlighting the salience of vocal cues in the context of recognizing deception.

Filled pauses, in particular, stand among a set of belief cues that listeners frequently associate with lie perception. Zuckerman, Koestner, and Driver (1981) found that speech hesitations were one of the cues reliably associated with judgment of lie-telling, whether it was in predicting behavior exhibited by subjects themselves or by others. This finding corroborates a parallel assessment of the actual cues that listeners use in detecting deception, where a meta-analysis of 33 studies again implicated filled pauses as an indicator of deception (Zuckerman, DePaulo, & Rosenthal, 1981).

The aforementioned studies thus outline a relationship between a speaker's manner of delivery and the listener's pragmatic interpretation of the speech; however, one limitation in the existing research is the reliance on off-line measures of judgment. While the time-course of pragmatic inferencing lies beyond the scope of those studies, off-line methods are consistent with an implicit assumption that in computing the meaning of an utterance, contributions made by pragmatic enrichment occur at a relatively late stage of processing. This is in line with traditional models of language comprehension, which frequently distinguish between the what and the how of a linguistic message. Under this view, listeners must first interpret the literal message content before consulting any social or contextual knowledge that may alter the global meaning of the utterance (Blank, 1988; Hamblin & Gibbs, 2003; Keysar, Barr, Balin, & Brauner, 2000; Lattner & Friederici, 2003). Pragmatic considerations, such as what might be intended by the utterance or whether the speaker might be lying, thus require additional (and correspondingly time-consuming) inferential processes before a contextually-appropriate meaning can be derived (e.g., Blank, 1988).

This temporal prioritization of the literal, however, has not gone unchallenged. A growing body of research emphasizes the primacy of the communicative aspect of linguistic exchange (Clark, 1996). Studies on figurative language processing suggest that people are not necessarily slower in interpreting the figurative meaning of an expression than the literal (e.g., Blasko & Connine, 1993; Gibbs, 1994; Giora, 1997; see Glucksberg, 2003 for a review). Using a cross-modal priming paradigm, Blasko and Connine (1993) measured the reaction times for participants making lexical decisions on words that were literally-related, metaphorically-related or unrelated to the vehicle of a metaphor that they heard, and showed that people access metaphorical meanings as quickly as the literal, even in the case of novel or unfamiliar metaphors such as "her thoughts were a boiling kettle". Given adequate contextual information, listeners can comprehend the intended meaning of non-literal utterances directly without first having to compute and reject a contextually-incompatible literal interpretation (Gibbs, 1994; cf. Pynte, Besson, Robichon, & Poli, 1996).

Findings such as these are inconsistent with traditional models, highlighting the problematic nature of the "literal-first" assumption. On the other hand, their conclusions converge with several studies examining the time course of pragmatic inferencing which have found immediate effects of various kinds of contextual constraints on listener interpretation (e.g., Nadig & Sedivy, 2002; Hanna, Tanenhaus, & Trueswell, 2003; Hanna & Tanenhaus, 2004; Van Berkum et al., 2008; although cf. Lattner & Friederici, 2003). For instance, results from event-related brain potentials (ERPs) reveal that listeners make inferences based on the speaker's voice, and what it suggests about age and gender. Evidence for these inferences emerges as early as 200-300 ms from the onset of a critical word, suggesting that pragmatic inferences are processed by the same early interpretation mechanisms that construct a content-based literal meaning (Van Berkum et al., 2008). Van Berkum et al. interpret their findings within a one-step model of language comprehension, where social context is incorporated immediately and alongside semantic content in the construction of an overall utterance meaning.

Eye-tracking studies on perspective-taking show that addressees are able to take into account common ground information available to both themselves and the speaker from the earliest moments of reference resolution (Hanna et al., 2003; Nadig & Sedivy, 2002). Hanna and Tanenhaus (2004) showed that listeners interpreting an ambiguous referent were sensitive to the contextual constraint of whether or not the speaker was physically able to reach the object, with effects emerging rapidly after the onset of the critical object name. Studies such as these show that listeners are able to rapidly extract and use pragmatic information based on the speaker's identity, and based on the situation that the utterance is produced in; two forms of context to which the listener has access prior to and during the entire utterance. Building on this, the experiments reported here are designed to establish whether listeners make rapid pragmatic inferences based on the moment-to-moment manner in which an utterance is spoken. Specifically, we investigate how a listener's on-line judgment of whether an utterance is a truth or lie varies with the voice-based cue of utterance fluency, exploiting the well-established finding which indicates a correlation between disfluencies and lie perception.

To date, the majority of research on the on-line comprehension of disfluencies has focused on the effect of disfluency on expectations relating to the semantic content of the message (J. E. Arnold et al., 2004; J. E. Arnold et al., 2007; see J. E. Arnold & Tanenhaus, 2011 for a review). For instance, J. E. Arnold et al. (2004) showed that eye movements of listeners told to manipulate objects on a visual display were biased toward either previously-mentioned or discourse-new objects, depending on whether the instruction contained a disfluency. Listeners hearing "Put the grapes above the candle. Now put the candle/camel…" were initially more likely to look at the candle, whereas disfluent instructions such as "Now put thee, uh, candle/camel..." elicited more looks toward the camel during the temporarily ambiguous onset ("ca..."). J. E. Arnold et al. (2007) later extended this finding to unfamiliar objects perceived as difficult to describe, by demonstrating that listeners hearing instructions such as "Click on thee, uh, green..." were initially more likely to look toward green squiggly objects than green everyday items such as ice cream cones. These studies provide evidence that listeners are sensitive to the manner of delivery when predicting the upcoming semantic content of the message; disfluencies create probabilistic expectations about which object a speaker is likely to refer to out of a given set of possible objects.

In a subsequent experiment, J. E. Arnold et al. showed that this semantic effect of disfluency is modulated by factors such as prior knowledge about the speaker. When they informed listeners that the speaker giving the instructions had object agnosia, they found that the tendency to fixate difficult-to-name objects was sharply reduced, highlighting the context-dependent nature of the disfluency bias (J. E. Arnold et al., 2007, Experiment 2). Barr and Seyfeddinipur (2010) observed a similar effect of disfluency, where listeners' content-based predictions were modulated by their inferences relating to the speaker's familiarity with the object. Using a mouse-tracking paradigm, they showed that disfluencies facilitated mouse movements toward a target when described by a speaker who had not seen the object before, but not during descriptions by a speaker familiar with the object. These studies show that various forms of contextual information can influence a listener's interpretation of disfluency when making on-line predictions about the message content. However, it remains unclear from the current literature how disfluencies may affect a listener's pragmatic hypotheses during the moment-to-moment processing of the linguistic message itself.

In the current study we aimed to investigate whether, and how, the presence of disfluency in an utterance influences listeners' pragmatic inferences relating to the speaker's truthfulness during on-line comprehension of the linguistic message (i.e., an inference about an alternative non-literal message). In the experiments reported here, we presented a speaker as being sometimes dishonest within a context framed as a lie detection task. Using a game in which listeners were led to assess whether or not a speaker was indicating the true location of a reward, we manipulated the speaker's manner of delivery (fluent vs. disfluent) to test how this influenced whether or not the listener believed the utterance. We made use of an eye- and mouse-tracking paradigm to establish the time course with which these pragmatic inferences are made. Together, the two measurements provided a picture of the focus of listeners' visual attention through their eye movements (Altmann & Kamide, 2007) and the continuous trajectory of decision-making through their mouse movements (Spivey & Dale, 2006). In Experiment 1, we used an utterance-initial filled pause to create the disfluent stimuli, in line with previous studies on manner of delivery (e.g., Brennan & Williams, 1995); in Experiment 2, we explored whether the location of disfluency matters by shifting the disfluency to an utterance-medial position. As we will show, the results demonstrate that listeners are influenced by both utterance-initial and utterance-medial disfluencies when making an implicit judgment on the veracity of the speaker's utterance. Eye- and mouse-tracking data also suggest that this inference is made during early moments of utterance comprehension—during fluent utterances, listeners were quickly biased toward the object referenced by the speaker, whereas with disfluent utterances they were biased toward the other object.

4.2.3 Experiment 1

Experiment 1 was designed to test whether a speaker's manner of delivery influences a listener's interpretation of an utterance as a truth or lie, as well as whether this judgment is formed during on-line language comprehension. We told participants that they were taking part in a lie detection study. The experiment was presented in the format of a computer game, with a series of opportunities for the participant to uncover hidden treasure behind one of two visible objects on the screen. Participants were told that the goal was to accumulate treasure by clicking on the object that concealed the treasure on each trial. Participants heard a recorded speaker reference which object to click on, but were informed beforehand that the speaker was a participant from an earlier experiment who had been instructed to lie half the time about the treasure's location. This served to establish an element of potential deception in the experiment, as well as to justify the presence of disfluency in the stimuli. We analyzed the participants' eye and mouse movements as well as their object clicks.

4.2.4 Method

Participants

21 self-reported native speakers of English took part in the experiment. Participants were all right-handed mouse users with normal or corrected-to-normal vision.

Materials

Visual stimuli comprised 120 line drawings from Snodgrass and Vanderwart (1980), presented in pairs across sixty trials (20 critical; 40 filler). For each pair presented in a visual display, we will use the term *referent* for the object that the speaker named as the object concealing the treasure; we will refer to the other object as the *distractor*. Critical referents and distractors were matched for ease of naming (H value < 1.0) and familiarity ($\geqslant 3.0$) to minimize participants' biases toward either object based on expectations relating to difficulty of description (cf. J. E. Arnold et al., 2007).[2] Care was also taken to ensure both objects did not start with the same sound on critical trials (cf. J. E. Arnold et al., 2004).

[2] H is a measure of name agreement for a given concept or image, where higher values correspond to greater difficulty experienced by speakers in naming the object. A description of its derivation can be found in Snodgrass and Vanderwart (1980), from which the H values and familiarity ratings reported here were taken.

Each referent was associated with a recording specifying the image as the object that the treasure was hidden behind. Critical utterances were either fluent (*The treasure is behind the ...*) or were preceded by an utterance-initial filled pause (*Um, the treasure is behind the ...*). The sentences were initially recorded in their entirety; a 400 ms filled pause from one of the disfluent utterances was then cross-spliced onto the start of each fluent statement to create a corresponding disfluent version. This ensured that participants were reacting to the same utterance (bar disfluency manipulation) in each condition, and to the same disfluency across all disfluent trials.

Table 4.1: Breakdown of filler stimuli and examples of each type.

Filler type	Manipulation	No. of utterances	Example
Fluent	None	20	The treasure is behind the giraffe.
Disfluent	Prolongation	3	The treasure is behind *thee...* mushroom.
	Repetition	4	The treasure is behind *the- the* roller skate.
	Filled pause (utterance-medial)	3	The treasure is behind the, *uh* rooster.
Other	Discourse marker	5	*Okay,* the treasure is behind the clothes peg.
	Modal	3	The treasure *could be* behind the balloon.
	Combination	2	*Right,* the treasure *might be* behind the caterpillar.

The 20 experimental referents were counterbalanced across two lists each containing ten fluent and ten disfluent utterances, such that each referent that occurred within a fluent utterance in the first list occurred within a disfluent utterance in the second. Each list included an additional 40 filler utterances, also naming a referent as the object concealing the treasure. To increase variability, half of these included one of various forms of disfluency, or a discourse manipulation such as a non-propositional sentence marker or a modal varying the speaker's commitment to the truth value of the statement. The reason for the filler manipulation was twofold: (a) to distract participants from the filled pause manipulation on critical trials, and (b) to create a set of utterances that closely approximated natural speech in order to reinforce the cover story which emphasised that sentences were unscripted. A summary of filler utterance types is provided in Table 4.1. The remaining 60 objects served as distractors. These were randomly paired with referents on each display, with no repetition of images across the experiment.

Procedure

The experiment was presented using OpenSesame version 2.9.4 (Mathôt et al., 2012) on a 21in. CRT monitor. Eye movements were monitored using an Eyelink 1000 Tower Mount system which tracked the right eye, sampling at 500Hz. Mouse coordinates were sampled at 50 Hz.

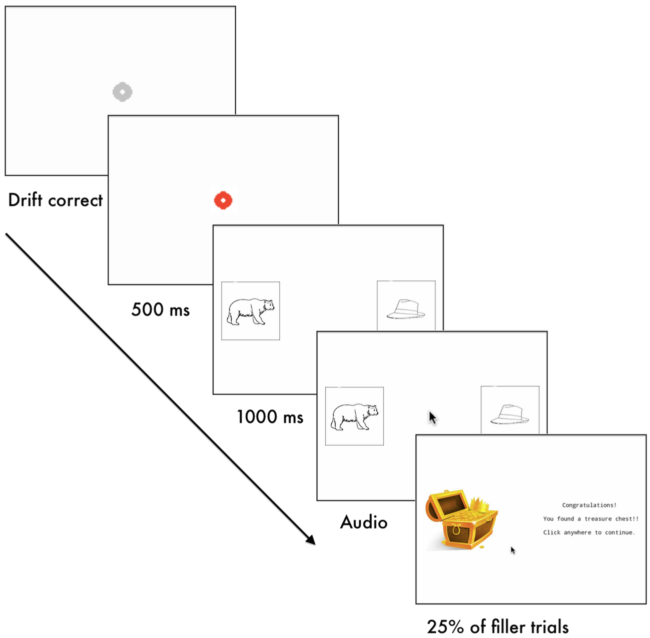


Figure 4.1: Timeline of a sample trial from the experiment

Fig. 4.1 presents a sample of a trial from the experiment. Between trials, participants saw a central gray fixation dot and underwent a manual drift correct to ensure accurate recordings from the eye tracker. After this, the fixation dot turned red for 500 ms to signify the start of the trial. The dot disappeared and was replaced by two images (a referent and a distractor), which were centered vertically and positioned horizontally left and right on the screen. Referents appeared on the left and right side equally often. Each pair of images appeared for a 1000 ms preview, after which the mouse pointer appeared at the center of the screen and playback of the audio stimulus began. Participants were instructed to click on the object which they believed concealed the treasure. Once a mouse click had been recorded on one of the two objects, the objects disappeared and the gray fixation dot appeared to begin the next trial, except in the case of fillers which included feedback (see below). Trials had an automatic time-out 5 seconds post-audio onset. If a click was not detected before this, participants saw a message telling them to respond more quickly.

To keep participants motivated, we informed them that the game contained a number of hidden "bonus" rounds which offered more treasure than the average trial. To simulate these rounds, 25% of filler trials were programmed to display a message that a treasure chest had been found regardless of which object was clicked on. This message appeared immediately following the detection of a mouse click on an object, and remained on screen until participants clicked again to begin the next trial. Additionally, participants were told that top scorers would be able to enter their name into the high score table, which appeared on the screen at the start of the experiment. A 5-trial practice session preceded the main experiment. One of the practice trials was always set to display the bonus treasure message. After the task, a post-experiment questionnaire was used to verify that none of the participants suspected that the audio stimuli comprised speech that had been scripted for purposes of the experiment.

For each experimental trial, we recorded eye and mouse movements as well as which object (referent or distractor) the participant clicked on.

4.2.5 Results

Trials on which a click was not recorded on either object were excluded from analysis (0.5% of experimental trials). Statistical analyses were carried out in R version 3.1.0 (R Core Team, 2017) using the lme4 package (Bates et al., 2014).

We modelled the binary outcome of final object click in a mixed effect logistic regression model to test for the main effect of manner of delivery (fluent/disfluent), including random intercepts and slopes for subject and item random effects. Eye-tracking records were averaged into 20 ms bins, each comprising ten samples, prior to analysis. Fixation data were coded in terms of region of interest (referent/distractor/none), and the proportion of fixations to each object out of the total sum of fixations was computed for each time bin. Mouse-tracking analysis only took into account the X coordinates. For each sample, the distance traveled by the mouse was computed by taking the difference between the X coordinates of the current and previous sample. The data were coded for direction of movement (toward referent/toward distractor) for each bin, and the cumulative distance participants had moved the mouse toward either object was calculated by summing over the distance traveled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). For each object, we then calculated a proportion-of-movement measure, defined as the distance traveled by the mouse pointer towards the given object, divided by the total distance traveled (regardless of X direction).

In all figures, the proportion of fixations or mouse movements to each object is plotted from onset of the referent (the point of disambiguation in the utterance) until 2000 ms post-onset, by which point participants had typically moved the mouse over one of the two objects. Model analyses for eye and mouse movements were conducted over a time window beginning from referent onset to 800 ms post-onset. This window was identified based on existing research which suggests listeners' eye movements establish reference around 400 to 800 ms after an object is mentioned (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Hanna et al., 2003; Tanenhaus, 2007). The window extends over the mean duration of critical referents (560 ms) and ends just after offset of the longest critical referent (776 ms), to minimise possibility of confounding results with any post-processing effects. Models were estimated using empirical logit regression (Barr, 2008), including time and manner of delivery (fluent/disfluent) as fixed effects, and random intercepts and random slopes for time in by-subjects and by-items analyses.

Final object click

The distribution of responses suggests an overall tendency to believe the speaker to be truthful rather than deceptive. Across the experiment, clicks on the referent were recorded in 58% of trials while clicks on the distractor were recorded in 42% of trials. Table 4.2 shows the breakdown of mouse clicks on each object in each condition. Participants were more likely to click on the referent following a fluent utterance than a disfluent one, $\beta = 2.30$, SE = 0.48, $p < .001$. Manner of delivery therefore appeared to influence perception of the speaker's truthfulness such that participants were more inclined to interpret a fluent utterance as truthful and a disfluent one as deceptive.

Table 4.2: Breakdown of mouse clicks recorded on each object (referent or distractor) by manner of delivery (fluent or disfluent) for Experiments 1 and 2. Values represent percentage of trials.

		Experiment 1	Experiment 2
Fluent	Referent	78	87
	Distractor	22	13
Disfluent	Referent	37	17
	Distractor	63	83

Eye movements

We calculated for each 20 ms time interval the proportion of fixations to each object, out of the total sum of fixations (including those to areas that constituted neither object). Fig. 4.2 shows the proportion of fixations to each object in each condition until 2000 ms post-noun onset. The time course of fixations reveals an effect that emerges rapidly in the fluent condition, with looks to the referent rising and diverging from the distractor beginning about 300 ms post-noun onset. In the disfluent condition, we initially see a rise in looks toward the referent, before an abrupt decrease around 600 ms post-onset and a corresponding increase in looks to the distractor. The initial rise in fixations to the referent likely reflects eye-movement sensitivity to establishing reference when an object is named (e.g., Eberhard et al., 1995); this explains the slight delay with which effects emerge during disfluent utterances relative to fluent.

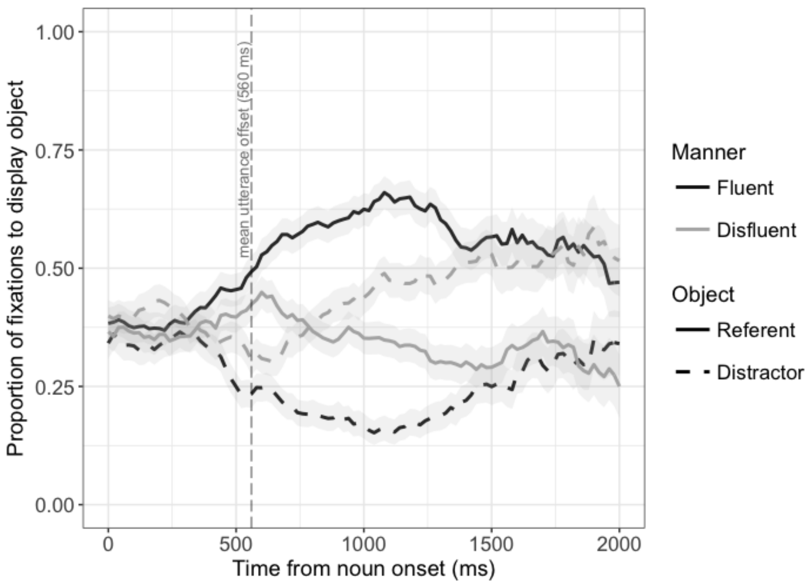


Figure 4.2: Eye-tracking results for Experiment 1: Proportion of fixations to each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ±1 standard error of the mean. Note that proportions do not sum to 1 because some fixations fell outside of either object.

To test how the effect of manner of delivery changed over time, we calculated the empirical logit of the fixation proportion to the referent for each 20 ms time bin over our 0 to 800 ms window of interest, averaging across all trials in each condition separately for subjects and for items. Model analysis revealed a reliable interaction between time and manner of delivery, β = 1.72, SE = 0.7, t = 2.47 by subjects; β = 1.01, SE = 0.39, t = 2.58 by items, reflecting the difference in listener commitment toward the object mentioned by the speaker dependent on manner of delivery.

Mouse movements

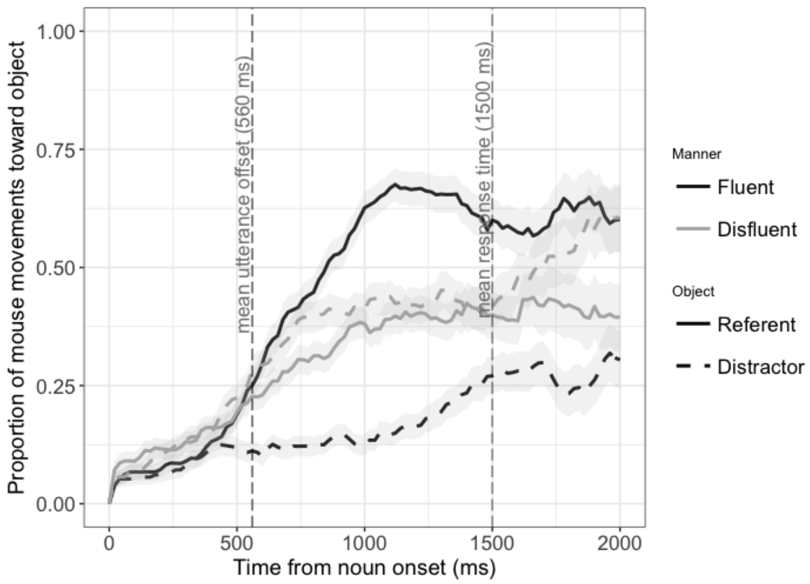


Figure 4.3: Mouse-tracking results for Experiment 1: Proportion of cumulative distance traveled toward each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ±1 standard error of the mean.

Mouse movements were scored according to whether they were in the direction of the referent or the distractor, and the absolute distance traveled (in pixels) was computed for each 20 ms time step. Fig. 4.3 shows the proportion of mouse movements (in terms of distance traveled) toward each object until 2000 ms post-noun onset. Examination of the total distance traveled by the mouse from the time of noun-onset to the point of response suggests an object bias dependent on manner of delivery. On fluent utterances, participants' cursors were more strongly attracted to the referent than the distractor, while disfluent utterances exhibited the opposite trend. This effect was confirmed by a model analysis on total mouse distance which revealed an interaction between object and manner of delivery, $\beta = 253.48$, SE = 87.67, $t = 2.89$ by-subjects; $\beta = 211.80$, SE = 89.15, $t = 2.38$ by-items.

The time course of mouse movements demonstrates that this difference emerged quickly post-noun onset. On fluent utterances, mouse movements toward the referent begin rising sharply and diverging from movements toward the distractor 400 ms post-onset. On disfluent utterances, movements reflect a small but distinct preference toward the distractor from around the same time, although a prominent divergence only appears about 1500 ms post-onset (see Fig. 4.3). Comparing movements toward the referent in both conditions though, it is clear that participants were more likely to move the mouse toward the referent during fluent utterances. This effect emerged 500 ms after onset of the noun, confirming that the disfluency affected listeners' interpretation during early moments of comprehension. Following the eye-tracking analysis, we computed the empirical logit of the proportion of distance traveled by the cursor toward the referent for each 20 ms time interval, calculated out of the total distance traveled in either direction. Results indicated a time by manner of delivery interaction for both subjects, $\beta = 7.47$, SE = 2.91, $t = 2.56$, and items, $\beta = 3.47$, SE = 1.50, $t = 2.30$. As with eye movements, participants' mouse movements appear rapidly sensitive to the manner of delivery such that fluent utterances are more likely to elicit movements toward objects named by the speaker compared to disfluent utterances.

4.2.6 Discussion

Experiment 1 established that listeners make pragmatic inferences about a speaker's truthfulness dependent on the manner of delivery of the utterance. Both eye- and mouse-tracking results indicate that this bias emerged over a time window corresponding to early moments of utterance processing. Although the time course suggests a slight temporal delay in mouse movements relative to the eye, this is unsurprising as it merely reflects the role of the visual system in the planning and execution of hand motor motions (cf. Land, Mennie, & Rusted, 1999). Accordingly, listeners' eye movements exhibited an initial sensitivity to the object being named during disfluent utterances, but a corresponding effect of mouse movements toward the referent was not observed. More importantly, the pattern of bias in the time-course of eye and mouse movements follow a similar trend. Fluent utterances yielded an increase in fixations and mouse movements to referent objects (i.e., participants were inclined to interpret the statement as truthful), while the opposite trend was observed for disfluent utterances. Taken together, these results indicate a clear effect of manner of delivery on the listener's interpretation of the speaker's truthfulness. Importantly, they suggest that there is rapid integration of pragmatic information during on-line comprehension, allowing participants to quickly assign a non-literal meaning to the speaker's mention of a referent. The speed with which they do this suggests that the disfluency has already been incorporated into their model of the speaker and has an immediate effect on their interpretation of the utterance.

4.2.7 Experiment 2

Experiment 2 was designed to test whether listeners' pragmatic inferences about speaker veracity are similarly sensitive to utterance-medial disfluencies. Previous production studies on disfluency suggest that utterance-initial disfluencies are linked to global, speech-planning issues. For example, Clark and Fox Tree (2002) observed that ums occurred most frequently at the start of intonation units, which they attributed to greater demands on the speech-planning mechanisms (cf. Swerts, 1998; Watanabe, 2002). Utterance-medial filled pauses, on the other hand, tend to be associated with localized, lexical access difficulties (e.g., Beattie & Butterworth, 1979). The design of comprehension studies to date have largely aligned with these accounts. Studies examining listeners' global, metacognitive perceptions of speech such as those relating to a speaker's state have mainly made use of utterance-initial filled pauses (e.g., Brennan & Williams, 1995), while those investigating listeners' expectations relating to the semantic content of the message have focused on the comprehension of utterance-medial pauses (e.g., J. E. Arnold et al., 2004). Hence, we tested whether the presence of disfluency mid-utterance affects listeners' global judgment of whether the speaker was lying or telling the truth. Based on findings from Experiment 1, if utterance-medial disfluencies similarly influence listeners' perception of deception, we would expect to see temporal evidence of this effect reflected in participants' early eye and mouse movements.

4.2.8 Method

Participants

22 participants took part in Experiment 2, all of whom fulfilled the same requirements as Experiment 1. None had participated in Experiment 1.

Materials

The same objects were used in Experiment 2 as in Experiment 1. Disfluent stimuli for both experiments were recorded during the same session to ensure consistent prosody with filler utterances. As with Experiment 1, the disfluent segment was cross-spliced from one of the disfluent utterances into each fluent utterance to create a disfluent counterpart. Disfluent utterances were characterized by an utterance-medial disfluent segment comprising a prolonged article followed by a filled pause (*The treasure is behind thee, uh …*). The 20 critical referents were counterbalanced across two lists each containing ten fluent and ten disfluent utterances. Filler utterances characterized by an utterance-medial filled pause in Experiment 1 were replaced accordingly for an utterance-initial filled pause. The rest of the filler utterances remained the same as Experiment 1.

Procedure

The experiment was identical in every respect to Experiment 1, with the only difference being the disfluency position in the recordings. A post-experiment questionnaire was again used to check whether any participants were sensitive to the experimental manipulations. Data from one subject who guessed that the speech was scripted was excluded from the analysis.

4.2.9 Results

We followed the same analysis procedures in Experiment 2 as in Experiment 1. Data from 4 experimental trials (0.9%) on which a click was not recorded on either object were excluded from the final dataset.

Final object click

The response distribution again suggests a global bias toward the object named by the speaker: Participants clicked on the referent in 57% of trials and on the distractor in 43% of trials. Table 4.2 shows the distribution of trials on which participants clicked on either object in each condition. As in Experiment 1, this was influenced by the speaker's manner of delivery, with participants more likely to click on the referent following a fluent utterance, $\beta = 4.06$, SE = 0.60, $p < .001$. This replication of results from Experiment 1 suggests that listeners are sensitive to utterance-medial disfluencies when making pragmatic inferences about a speaker's truthfulness. To test whether there were differences in sensitivity to utterance-initial and utterance-medial disfluencies, we compared final object clicks for the subset of disfluent trials in both experiments. Participants were less likely to click on the referent following utterance-medial disfluencies, $\beta = -1.52$, SE = 0.52, $p < .005$. We return to this effect in the General Discussion.

Eye movements

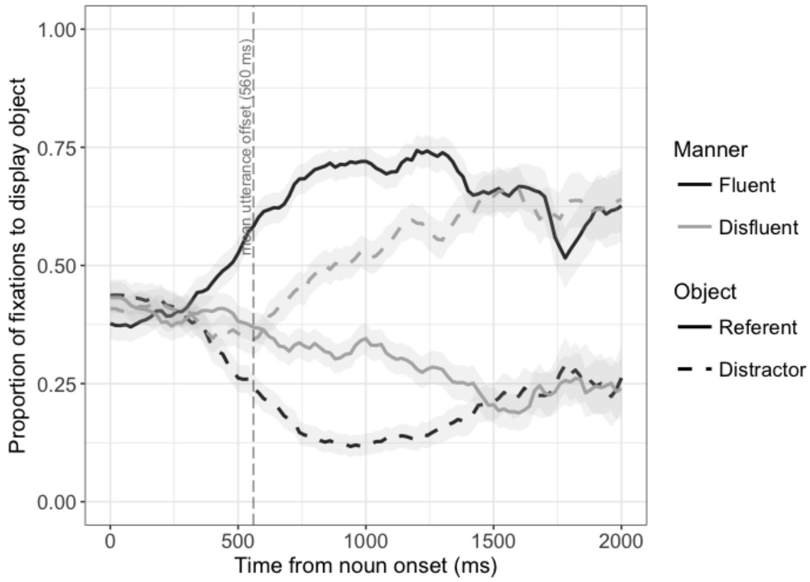


Figure 4.4: Eye-tracking results for Experiment 2: Proportion of fixations to each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ±1 standard error of the mean. Note that proportions do not sum to 1 because some fixations fell outside of either object.

Fig. 4.4 shows the proportion of fixations to each object in each condition until 2000 ms post-noun onset. The effect of manner of delivery is reflected in eye movements as the speech unfolded over time. During fluent utterances, fixations to the referent rose quickly, accompanied by a corresponding decrease in looks to the distractor. Conversely, disfluent utterances yielded the reverse pattern. The time course of events is comparable to that observed in Experiment 1, with a fixation bias emerging 300 ms after onset of the referent in the fluent condition, and around 600 ms post-onset in the disfluent condition. This effect on looks to the referent over time was reflected in a significant interaction between time and manner of delivery, for both subjects and items analyses on the time window $0 - 800$ ms post-noun onset, $\beta = 3.82$, SE $= 1.33$, $t = 2.86$, and $\beta = 2.96$, SE $= 0.95$, $t = 5.01$. A comparison of eye movements during disfluent utterances across the two experiments found no effect of disfluency location, $\beta = -0.26$, SE $= 0.32$, $t = -0.81$.

Mouse movements

Mouse-tracking results reveal a contrast between fluent and disfluent utterances which corroborates the eye-tracking. Analysis of the distance traveled toward each object as a proportion of the total distance participants moved the mouse suggests a bias due to manner of delivery: on fluent utterances, participants' cursors were overwhelmingly more likely to be attracted to the referent, while disfluent utterances were characterized by more movements toward the distractor. This difference was confirmed by a significant interaction between object and manner of delivery, $\beta = 475.95$, SE $= 96.95$, $t = 4.91$ by subjects; $\beta = 377.35$, SE $= 70.33$, $t = 5.37$ by items.

Fig. 4.5 shows the proportion of mouse movements (in terms of distance traveled) toward each object until 2000 ms post-noun onset. Time course examination demonstrates that the object bias emerged rapidly after onset of the disambiguating noun, with a referent-bias beginning 300 ms post-onset during fluent utterances, and a distractor-bias beginning 200 ms during disfluent ones, showing that participants' mouse movements were quickly sensitive to the speaker's manner of delivery in their decision of which object to click on. As in Experiment 1, model analysis taking the empirical logit of the proportion of distance traveled toward

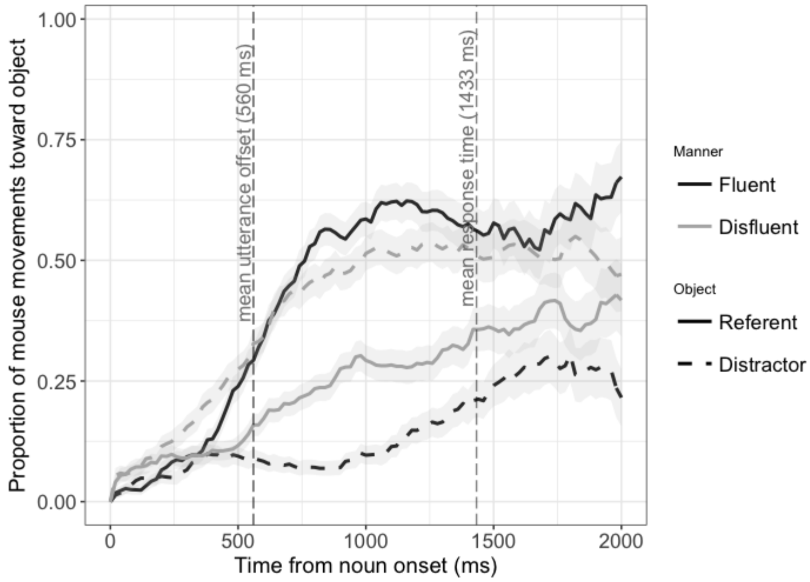


Figure 4.5: Mouse-tracking results for Experiment 2: Proportion of cumulative distance traveled toward each object (referent or distractor) in each condition (fluent or disfluent), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ± 1 standard error of the mean.

the referent as the dependent variable yielded a significant interaction between time and manner of delivery, both by-subjects, $\beta = 11.04$, SE = 2.69, $t = 4.10$, and by-items, $\beta = 6.73$, SE = 2.82, $t = 2.39$. As with the eye movements, there was no effect of disfluency location on mouse movements when comparing Experiments 1 and 2, $\beta = 2.22$, SE = 2.90, $t = 0.76$.

Post-hoc analyses

We conducted two subsidiary analyses to rule out potential counterexplanations of our findings. The first examined whether participants who knew they might be deceived might become sensitive to the distributions of fluent/disfluent utterances, using them as discriminative cues due to the nature of the task. The second considered whether disfluencies were understood as a general signal of uncertainty, rather than as a specific signal of deception.[3] To maximise power, we present analyses from data pooled across Experiments 1 and 2, although we note that the findings from analyses of individual experiments do not differ.

In order to establish whether participants learned to discriminate the disfluent signal, we compared the trials encountered during the first third of each experiment to those encountered during the last third. There was no effect of trial subset for either eye movements, $\beta = -0.09$, SE = 0.30, $t = -0.29$, or mouse movements, $\beta = -0.17$, SE = 0.48, $t = -0.36$, confirming that the results were unlikely to be a learned response due to participants picking up on experimental contingencies over time.

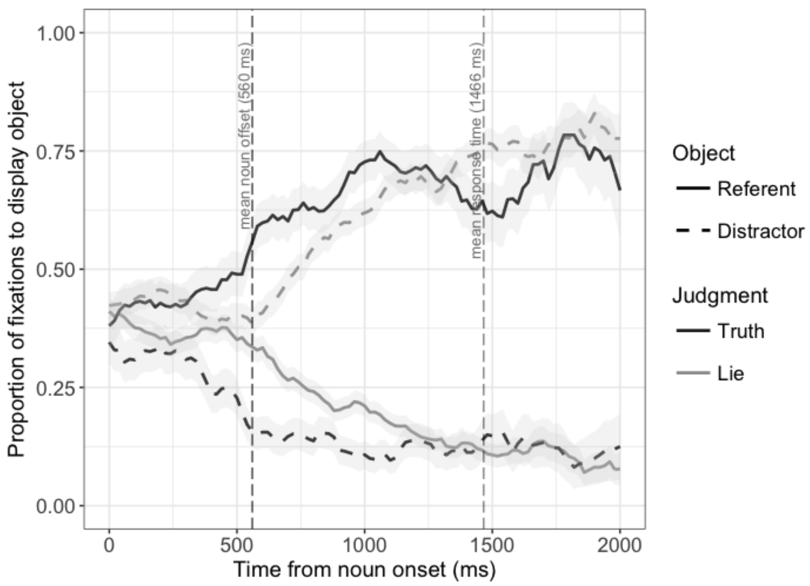


Figure 4.6: Eye-tracking results for subset of disfluent trials from Experiments 1 and 2: Proportion of fixations to each object (referent or distractor) broken down by response judgment (truth or lie), from 0 to 2000 ms post-noun onset, calculated out of the total sum of fixations for each 20 ms time bin. Shaded areas represent ± 1 standard error of the mean.

[3]We thank a reviewer of an earlier version of this paper for these suggestions.

To test whether disfluencies were perceived as general signals of uncertainty (cf. Brennan & Williams, 1995), we analyzed only those items containing a disfluency, comparing trials that resulted in a judgment of truthfulness versus deception. Figs. 4.6 and 4.7 show that participants' eye and mouse movements demonstrate an early object bias contingent on the final response judgment (based on the object that they ultimately clicked on for that trial); trials which resulted in a judgment of truthfulness saw a referent bias, while those resulting in a judgment of deception saw a distractor bias. This difference was reflected in a time by object interaction for both fixations, $\beta = 2.23$, SE = 0.29, $t = 7.61$, and mouse movements, $\beta = 12.32$, SE = 3.02, $t = 4.08$, confirming that the effect of disfluency on listeners' inferences about the speaker's truthfulness occurred during the initial stages of comprehension.

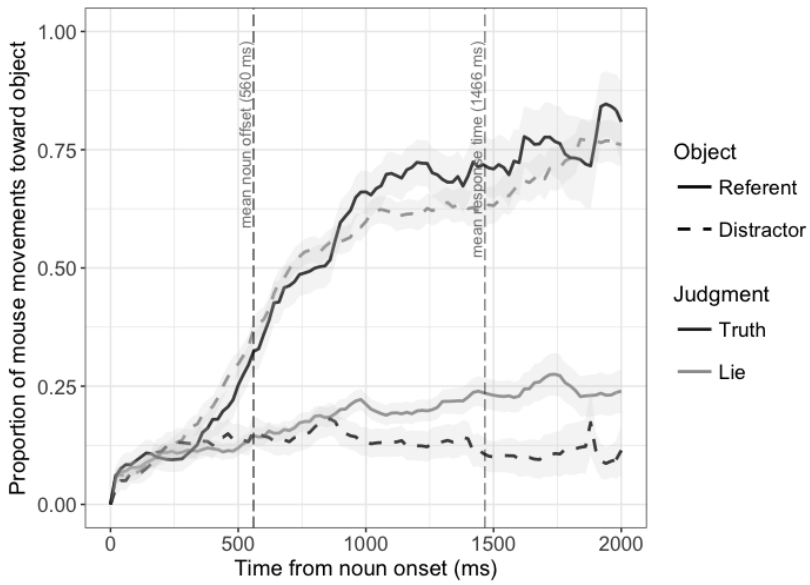


Figure 4.7: Mouse-tracking results for subset of disfluent trials from Experiments 1 and 2: Proportion of cumulative distance traveled toward each object (referent or distractor) broken down by response judgment (truth or lie), from 0 to 2000 ms post-noun onset. Proportions were calculated out of the total cumulative distance participants moved the mouse from noun-onset until that time bin. Shaded areas represent ±1 standard error of the mean.

4.2.10 General Discussion

The distribution of responses in both experiments indicates an overall tendency to believe the speaker was being truthful rather than deceptive. While this result differs from what we might expect by chance, it follows a trend observed in previous studies on proposition verification, which provide evidence to suggest an *a priori* prejudice in listeners toward believing something to be true (e.g., Barres & Johnson-Laird, 2003; McKinstry, Dale, & Spivey, 2008). These studies show that when asked to assess the truth value of a statement, participants demonstrate greater difficulty in evaluating a statement as false, suggesting an expectation regarding at least a partial adherence to the Gricean maxim of quality on the speaker's part. In the current study, we observed that this pre-existing "truth-bias" still appears to bear weight even when listeners have been told to expect a potentially non-Gricean, uncooperative speaker. Notably, this trend also aligns with findings from actual lie detection studies, which point at a general tendency toward perceiving speakers to be truthful rather than deceptive. For example, Vrij et al. (2000) found a mean of 61.5% truth judgments across nine studies analyzing listeners' truth-lie discrimination accuracy. This lends face value to the suggestion that the behaviour of participants in the current study was comparable to that of judges discriminating between authentic truths and lies.

Extending previous findings on the effect of manner of delivery, our results show that the way in which a message is conveyed affects a listener's implicit judgment of the speaker's truthfulness. Specifically, fluent utterances bias listeners toward perceiving a speaker to be truthful, as evidenced by increased fixations on, and mouse movements towards, the object named by the speaker, while disfluency biases listeners toward expecting the speaker to lie. This is consistent with previous studies on deception, which show that listeners believe liars to be more disfluent than truth-tellers. Together, the two experiments establish that listeners' pragmatic interpretations are affected very quickly by manner of delivery, whether disfluency occurs utterance-initially or mid-utterance.

Utterance-medial disfluencies resulted in fewer clicks on the named referent than did utterance-initial disfluencies. With the caveat that this is a between-experiment comparison, this difference appears difficult to reconcile with the production-based view that lies are planned at the level of the message, and that any difficulty associated with lying should therefore be best evidenced by utterance-initial disfluencies. This highlights the complexities associated with extending production accounts of disfluency to listener comprehension. While speakers may be disfluent in ways that are predictable, listeners' sensitivities to disfluencies might depend on multiple factors, such as their ability to model the mind of the speaker, or the contexts in which these cues arise. For instance, a phrase-medial filled pause may arise as a consequence of a speaker's search for a word; however, in the context of anticipating a potentially dishonest speaker, the listener may interpret the disfluency as a pragmatic cue to deception. Such a disjunct between speakers' productions and listeners' perceptions of disfluencies has been recently shown in the context of a competitive game involving deception (Loy, Rohde, & Corley, 2016b).

Contrary to predictions made by traditional models of language comprehension, temporal information from the eye- and mouse-tracking provides little evidence to suggest that pragmatic inferences are relegated to a later, post-literal stage of processing. Rather, our results support recent research showing that pragmatic information is extracted and used as an immediate constraint on utterance processing (e.g., Hanna & Tanenhaus, 2004; Nadig & Sedivy, 2002). Recall that in order to isolate effects during on-line comprehension, our time course analyses were conducted over a window beginning from noun-onset and ending 800 ms post-onset. Given that the average duration of critical nouns in the experiment was 560 ms (range: $413 - 776$), when taking into account the 200 ms it typically takes to program an eye-movement (Matin et al., 1993), it is clear that participants' eye and mouse movements began to exhibit this bias as the critical linguistic input unfolded. Thus, pragmatic judgments about the veracity of the speaker's utterance were made on-line, and modulated immediately by the presence or absence of a paralinguistic cue.

Building on previous studies which show that disfluency affects listeners' on-line expectations about the literal message content, our findings demonstrate that disfluency modulates listeners' pragmatic hypotheses about the speaker's intentions simultaneously alongside integration of lexical, semantic information. The speed with which these hypotheses unfold is consistent with the disfluency bias observed by previous studies, delineating an immediacy associated with judgment biases triggered by paralinguistic information such as filled pauses. Within the framework of detecting deception, this immediacy is also compatible with longstanding stereotypes of deceit that people seem to retain, despite repeated findings that perceivers tend to perform at or near chance level at overt lie detection tasks (e.g., ten Brinke, Stimson, & Carney, 2014). Given the ingrained nature of these stereotypes, it is perhaps less surprising that such biases are hard to overcome, whether in the context of judging deception or some other pragmatic aspect of the utterance in question.

One question arising from the results here is why the distractor-bias observed in the disfluent condition is smaller than the referent-bias in the fluent condition. One possible explanation could be due to the nature of the experimental stimuli, which resulted in fluency being made salient in the presence of disfluent utterances. Consequently, participants may have made their judgment based on the speaker's relative fluency (rather than simply treating disfluency as a systematic indicator of deception), resulting in what could be explained as an effect of fluency in response to disfluency. However, because participants had a pre-existing bias toward expecting the speaker to be truthful, this created a baseline preference for the object identified by the speaker. This in turn had to be overcome in order for participants to choose the distractor object during utterances perceived as lies, resulting in a greater referent-bias during fluent utterances compared to the corresponding distractor-bias during disfluent utterances. This raises the question of whether there are probabilistic cues to deception (such as disfluency in speech) that listeners implicitly rely on to recognize a deceptive speaker, or whether the process is in fact modulated by contextual reasoning (such as perception of an unexpected change in the speaker's manner of delivery). If the latter were the case, a disfluent utterance may be less likely to be perceived as a lie when produced by a hesitant speaker compared to one that was consistently fluent. Exploring how listeners integrate various sources of information when making pragmatic inferences would be a useful avenue for future research.

To our knowledge, this is the first study to integrate eye- and mouse-tracking measures in a visual world paradigm. It would appear that they corroborate each other, with the eyes tending to fixate the target that listeners are moving the mouse towards. In addition to validating conclusions from previous studies that mouse coordinates can serve as a temporally sensitive index of language processing comparable with eye movements (Farmer et al., 2007; Spivey et al., 2005), this opens up opportunities for further studies seeking to employ a mouse-tracking only methodology (e.g., in the case of certain clinical or developmental populations, which can sometimes pose challenges when eye-tracking; cf. Sasson & Elison, 2012), or even studies combining the two to more fully understand the perceptual-motor processes underlying spoken language comprehension.

The current results show that listeners make use of information in a disfluency to draw pragmatic inferences about a speaker's truthfulness, and more importantly, that this information is brought to bear during initial stages of the comprehension process. These results are relevant not only to comprehension in contexts which involve deliberate deception, but also in the many contexts in which a listener believes a speaker to be cooperative but must still take care to consider the status of the literal message—is it true? Is additional meaning intended? (e.g., "your haircut looks, um, great"; "the house for sale is, um, cozy"). In this way, we propose that the immediacy of listeners' pragmatic use of disfluency in evaluating an utterance is applicable broadly in language comprehension.

4.3 Chapter discussion

This chapter set out to explore how a speaker's manner of delivery influences a listener's perception of deception. Modifying the paradigm we used in Chapter 3, we created a task in which listeners made an implicit judgement of a speaker's veracity by choosing between two objects based on the speaker's (potentially untruthful) description of which object to pick. Our results demonstrate an effect of manner: Fluent utterances biased listeners toward believing the speaker, while disfluent utterances biased listeners in the opposite direction. Importantly, these effects emerged during the initial stages of comprehension, outlining the immediacy with which such cues are brought to bear in the process of reasoning about a speaker's intent to deceive.

The finding that disfluency biased listeners toward perceiving an utterance as a lie suggests that listeners pragmatically reason about the cause of a speaker's disfluencies within a model that links lying with cognitive effort, and cognitive effort to disfluency. Such a model is best captured by the cognitive hypothesis to deception, which attributes cues such as increased speech disturbances to the cognitive load associated with lying (Vrij et al., 2008). This association is consistent with evidence from some lie production studies that indicate an increase in disfluencies, in particular filled pauses, as a result of the cognitive demands of lying (Vrij, 2000; Vrij & Heaven, 1999; although see Arciuli, Mallard, & Villar, 2010). More generally, they also align with evidence from research outside of deception to suggest a general association between disfluency and perceived difficulty in speech production (J. E. Arnold et al., 2007; Barr, 2001).

Together, results from the two experiments in this chapter reveal that listeners' veracity judgements are sensitive to the speaker's manner, regardless of where disfluency occurs within an utterance. This finding is difficult to reconcile with theories of disfluency production, which suggest that planning difficulties, such as having to formulate a lie, are more likely to result in disfluencies at the start of an utterance rather than in the middle. Within the context of perceiving deception, this may point to the stereotypes of deceit that listeners hold, which highlight a strong association between disfluency and lying (Zuckerman, Koestner, & Driver, 1981). On a broader level, it also suggests that listeners may not interpret disfluencies within the same model that predicts when a speaker may be disfluent. This somewhat undermines a communicative account by which disfluencies are planned and produced as listener-oriented signals (Clark, 1996; Clark & Fox Tree, 2002), raising questions about the extent to which listeners and speakers align on the role of such verbalisations in conversation. This is a topic we put aside for now but return to in Chapter 6.

The results from this chapter further inform the questions that this thesis is concerned with on two levels. First, building on our results from Chapter 3, we provide further evidence that listeners' pragmatic inferences about an utterance unfold from the earliest stages of comprehension. As with our findings from Experiment 3.1, these results are incompatible with the standard pragmatic model, which relegates pragmatic understanding to a later, post-literal stage of comprehension. Instead, we show that the interpretation derived by listeners is shaped by the pragmatics of the communication from the earliest moments of meaning construction. Second, the results also align with findings from the previous chapter to show that the manner in which a message is conveyed has a direct and rapid impact on a listener's pragmatic interpretation. This perceptual relevance goes beyond contexts where meaning may simply be ambiguous between two (valid) forms, to cases where meaning may be the opposite of what the speaker actually says. Thus, the effect of manner is not limited to a single aspect of pragmatic comprehension; rather, listeners appear able to make rapid use of disfluencies in flexible ways to guide their pragmatic hypotheses about the meaning of an utterance. These results contribute to our general understanding of the process by which listeners' pragmatic inferences are shaped by a cue in the speaker's manner of delivery. In the next chapter, we examine whether this effect of manner remains relevant in contexts where listeners are presented with other sources of pragmatic information which may influence their perception of deception.

Chapter 5

Speaker-specific expectations on the effect of manner

In the previous chapter, we established that listeners were rapidly sensitive to a speaker's manner of delivery when making inferences about the speaker's veracity. Fluent utterances biased listeners toward perceiving the speaker to be truthful, while disfluent utterances biased them toward inferring that the speaker was lying. Listeners' eye movements and mouse trajectories revealed that these veracity judgements were made on-line, and modulated immediately by the presence of disfluency in the utterance. In the current chapter, we address the question of whether this disfluency-lying bias in listeners is modulated by additional sources of information, ones which create speaker-specific expectations that could potentially provide alternative explanations for the disfluency.

This chapter reports two experiments designed with the aim of exploring the effect of speaker-specific expectations on listeners' disfluency-lying bias. The first experiment focused on expectations based on a speaker's perceived veracity. We modified the paradigm used in Chapter 4 to include two speakers, varying in terms of veracity. The second experiment focused on expectations based on a speaker's accent, which served as a proxy for production difficulty. Here, we modified the paradigm to incorporate a native and a non-native speaker of English. These changes allowed us to test whether listeners' disfluency-dependent inferences about utterance veracity differed across the two speakers in each experiment. In each experiment, we found that even though listeners developed distinct expectations about each speaker, ultimately their disfluency-lie bias was robust against these expectations. This suggests that in contexts in which listeners entertain possible deception, disfluency may simply be a "universal" cue that listeners associate with lying regardless of the speaker.

5.1 Speaker-specific expectations can influence comprehension

While limited work to date has examined how listeners' expectations about a speaker may affect their perception of deception, speaker-specific expectations are known to impact other aspects of comprehension. For example, stereotypes based on a speaker's gender (K. Johnson, Strand, & D'Imperio, 1999; Strand, 1999) or their perceived ethnicity (Staum Casasanto, 2008) can affect how listeners perceive ambiguous phonetic input. Top-down knowledge about a speaker, such as information about their occupation, has also been shown to create expectations that influence a listener's interpretation. In particular, this type of expectation can affect pragmatic comprehension, where listeners often form inferences beyond a literal interpretation of the utterance. Metaphors such as "children are precious gems", for instance, are judged as more sarcastic when associated with speakers from high-irony occupations (e.g., comedian, movie critic), compared to speakers from high-metaphor occupations (e.g., teacher, artist; Katz & Pexman, 1997). Mol, Kuhlen, van der Steen, and Obbens (2013) demonstrate the role of occupation-dependent expectations on a different aspect of pragmatic comprehension: that relating to inferences about a speaker's certainty. Based on earlier findings that listeners' judgements of a speaker's certainty varied with the speaker's manner of delivery (Brennan & Williams, 1995; Swerts & Krahmer, 2005), Mol et al. hypothesised that top-down information about the speaker's expertise would influence listeners' certainty judgements. In line with this, they observed that statements about gardening were rated as more certainly correct when listeners were told the speaker was a gardener, compared to when they were told the same speaker was a Dutch teacher. These findings suggest that global beliefs about a speaker are encoded during the comprehension process to shape the overall interpretation derived by a listener.

In addition to influencing their global, off-line judgements, speaker-specific expectations have been shown to guide listeners' on-line comprehension processes. Metzing and Brennan (2003) demonstrate that expectations based on conceptual pacts between conversation partners (cf. Brennan & Clark, 1996) can influence a listener's comprehension of a speaker's referring expression. When a speaker who had previously used a label for an item (e.g., "the shiny cylinder") switched to a new expression (e.g., "the silver pipe"), listeners' looks to the relevant target were delayed, and drawn, instead, to other objects in the display. In contrast, the same delay was not observed when the new expression was uttered by a different speaker who had not used either label before.

Grodner and Sedivy (2011) extended Metzing and Brennan's findings by showing that speaker-specific expectations can constrain how a listener perceives the *type* of referring expression used by a speaker. Grodner and Sedivy had listeners manipulate objects in displays following instructions that contained a modified noun phrase (e.g., "Pick up the *tall* cup"). The displays featured a contrast set of items (e.g., a tall cup and a short cup), a competitor that shared the modifier property of the target (e.g., a tall pitcher), and a distractor. Consistent with previous work (Sedivy, 2001; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), Grodner and Sedivy observed that listeners interpret such modified noun phrases contrastively, directing their gaze toward the target over the competitor following onset of the relevant modifier. Importantly, they additionally showed that this effect was speaker-specific: No such bias was found when the same contrastive modifier was used by an unreliable speaker who frequently uttered such descriptions infelicitously. Grodner and Sedivy suggest that referential contrast effects are defeasible, and occur via a process by which listeners take into account the characteristics particular to a given speaker when inferring whether to draw the contrast meaning or not.

The collective message behind these findings is clear: Listeners interpret what is said differently depending on who says it (cf. Van Berkum et al., 2008). An extension of this research then is the question of whether listeners' interpretation of *how* something is said depends on who says it. This question has been addressed within the literature on prediction. An emerging body of research shows that speaker-specific expectations can influence listeners' interpretations of a speaker's disfluencies in the context of predicting upcoming content. J. E. Arnold et al. (2007) had listeners follow instructions to click on objects in displays that featured familiar, everyday items (e.g., ice cream cones) alongside unfamiliar, difficult-to-name items (e.g., abstract squiggles). Disfluent instructions were found to bias listeners' eye movements toward difficult items prior to the speaker naming the item; however, this bias disappeared when listeners were told the speaker had object agnosia, and hence might be presumed to struggle with naming familiar items.

In a similar vein, Barr and Seyfeddinipur (2010) demonstrated that listener-derived inferences about a speaker's familiarity with objects can modulate their disfluency-dependent biases during prediction. When selecting between images that were new (previously unseen) or familiar (having appeared in previous trials) to a speaker based on the speaker's description, listeners' mouse movements were more strongly attracted toward new images when the description was preceded by an *um*. This finding is in line with earlier studies which show that listeners focus on new objects following a disfluent instruction (J. E. Arnold et al., 2004; Barr, 2001). Importantly, Barr and Seyfeddinipur found that this effect of disfluency depended on the speaker: When descriptions were produced by a speaker who had previously seen one of the images, disfluency biased listeners toward the other (new) image; however, when a different speaker who was new to both images was disfluent, listeners' mouse movements showed no bias toward either image. These findings suggest that the degree to which disfluency guides listeners' predictive processes varies depending on their expectations about the speaker's cognitive state. When led to believe a speaker might have difficulty naming all the objects within the context, listeners cease to rely on the speaker's disfluencies to predict upcoming content.

Expectations based on a speaker's accent have also been implicated as a factor in listeners' interpretation of disfluent or erroneous speech. Bosker, Quené, Sanders, and de Jong (2014) observed that listeners following instructions to click on objects in a display exhibited anticipatory eye movements toward low-frequency lexical items (e.g., sewing machine) over high-frequency ones (e.g., shoe) following disfluent instructions (cf. Hartsuiker & Notebaert, 2010; Schnadt & Corley, 2006). However, this disfluency bias was attenuated when the instructions were produced by a non-native speaker, who would presumably struggle with a greater degree of cognitive difficulty in speech production (cf. Segalowitz, 2010). This finding suggests that listeners may adjust their disfluency-driven predictions about upcoming content based on their accent-dependent expectations about the speaker's cognitive demands. Their results are also in line with previous research demonstrating a degree of listener flexibility in comprehending pragmatically infelicitous sentences (e.g., "The mother gave the candle the daughter") when produced by a non-native speaker (Gibson et al., 2017; Hanulíková, Van Alphen, Van Goch, & Weber, 2012).

Taken together, these studies provide evidence that listeners' interpretations based on a speaker's manner of delivery are dependent on the specific speaker. Crucially, this appears to be modulated via a mechanism by which disfluency is associated with speaker difficulty, whether from describing unfamiliar topics or from speaking in a non-native language. This raises the question of whether listeners may similarly reason about speaker-specific difficulty in the context of perceiving deception, and how this might consequently affect their perception of the speaker's disfluencies.

One reason for expecting that listeners' disfluency-lying bias might be specific to the speaker is based on the *cognitive hypothesis* to deception, which links cognitive load (and by extension disfluency) to deception (Vrij & Granhag, 2012). Under this hypothesis, lying is taken to be a cognitively demanding task due to a variety of reasons, such as for example, having to maintain consistency between past lies and current evidence (Vrij, Ennis, Farman, & Mann, 2010). In support of this, lie production studies show that lying results in cues associated with greater cognitive load, notably an increased production of disfluencies (Vrij & Heaven, 1999). Moreover, listeners appear to subscribe to this model of behaviour: Lie perception studies show that people associate disfluencies with lying, whether in predicting behaviour exhibited by themselves or by others (Zuckerman, Koestner, & Driver, 1981). If listeners model the cognitive behaviour of speakers, they might reason that the production of manner-based cues associated with lying might vary depending on the cognitive load of the speaker at hand. For instance, a speaker known to often deceive might be expected to encounter less difficulty lying, and therefore be less likely to produce a disfluency when they lie (e.g., prisoners associate fewer cue behaviours with deception compared to police officers and prison guards; Vrij & Semin, 1996). Expectations based on cognitive load might also be implicated via the speaker's accent. Non-native speech might imply greater demand on various aspects of the production system (Segalowitz, 2010), potentially reducing the likelihood that disfluency is associated with lying.

The experiments in this chapter focus on these two sources of speaker-specific expectations to test whether listeners' tendency to attribute disfluency to lying depends on the given speaker. By modifying the paradigm from the previous chapter to include two speakers, we create a context by which listeners can draw on specific beliefs about each speaker to potentially influence their interpretation of each speaker's utterances.

5.2 Experiment 5.1

This experiment was designed to test whether listeners' disfluency-lying bias is modulated by their expectations based on a speaker's perceived veracity. The basic paradigm was similar to that of Chapter 4, in which participants attempted to uncover hidden treasure based on a description of the treasure's location by a potentially deceptive speaker. The experiment was modified such that the utterances were produced by one of two speakers, who differed in terms of relative veracity. Participants were told these utterances were taken from an earlier experiment, in which speakers produced a mix of true and false utterances about the treasure's location.

The experiment included a training phase, with the aim of establishing a bias toward expecting one speaker to lie consistently more than the other. Hence, this phase included feedback on each trial to reveal that one speaker told the truth on the majority of trials (22 out of 30), while the other lied on the majority of trials (same frequency). Following the training phase, participants went on to the experimental phase, which omitted feedback and introduced the element of disfluency to speakers' utterances. Participants' eye and mouse movements were recorded on each trial, as well as the final object that they clicked on.

5.2.1 Method

Participants

Thirty-two self-reported native British English speakers took part in the experiment. None had participated in the experiments reported in Chapter 4. Participants were all right-handed mouse users with normal or corrected-to-normal vision and no speech or hearing disorder. All participants provided informed consent in accordance with the university's Psychology Research Ethics Committee guidelines (Ref. No.: 137-1516-1).

Materials and design

Visual stimuli comprised 240 black and white line drawings from the Snodgrass and Vanderwart (1980) dataset. Half of these were used in the training phase and the other half in the experimental phase. The images were presented in pairs across 60 trials in each phase.

Each image pair was presented alongside a recording in which a speaker named one image as the object concealing the treasure (the referent). The other object served as a distractor. The recorded utterances were produced by two native British English speakers (both male), recorded individually in a sound-attenuated booth. Utterances were normalised to have the same mean acoustic intensity. The two speakers' recordings were also pitch manipulated by a factor of 0.9 and 1.1 respectively to magnify the difference between the voices.

Two speaker identities were created—Fred the pharmacist and John the journalist. Each speaker identity was associated with a Lego figure avatar corresponding to his occupation (see Fig. 5.1). For every experimental session, each speaker's voice was associated with one identity (Fred/pharmacist or John/journalist; counterbalanced across participants).

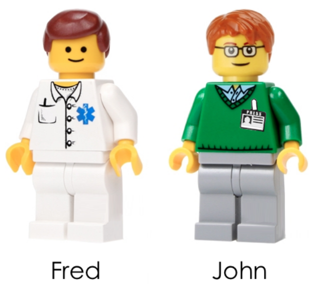


Figure 5.1: Experiment 5.1: Speaker identities and their respective avatars

The training phase consisted of 60 trials, comprising 30 utterances produced by each speaker. These were presented in a random order across the training phase. Utterances were all fluent forms that referenced one object as the treasure's location (e.g., *The treasure is behind the hat*). This phase was designed to present one speaker as more truthful than the other: Feedback after each trial revealed one speaker as having told the truth 73% of the time (22 out of thirty trials), and the other 27% of the time. We refer to these relative degrees of veracity as *truthful* and *dishonest* respectively. Training referents were drawn from a set of 60 objects, counterbalanced across two lists such that each object referenced by the truthful speaker in one list was referenced by the dishonest speaker in the second. The identity of each speaker (Fred or John) as well as their relative veracity (truthful or dishonest) was also counterbalanced across participants, creating a total of eight lists in the training phase.[1]

The experimental phase consisted of 60 trials, with 30 utterances produced by each speaker (10 critical; 20 filler), presented in a random order. Critical utterances were either fluent or contained an utterance-medial disfluency comprising a prolonged article and a filled pause (*The treasure is behind thee, uh hat*). All utterances were initially recorded in their entirety; the disfluency segment (*thee, uh*) from each disfluent utterance was then cross-spliced into its corresponding fluent utterance to create a disfluent counterpart. This ensured that participants were reacting to the same utterance (bar disfluency manipulation) for each speaker across fluent and disfluent conditions. Critical referents were drawn from a set of 20 objects, counterbalanced across four lists in a 2 (fluent/disfluent) x 2 (truthful speaker/dishonest speaker) design. As with the training phase, the identity of

[1]Unlike earlier studies which directly investigated the effect of occupational stereotypes on comprehension (e.g., Katz & Pexman, 1997; Mol et al., 2013), we introduced speaker occupations merely for the purpose of providing an additional cue which listeners could rely on to distinguish between the two speakers. Hence, this was counterbalanced across participants to avoid potential confounds based on occupational stereotypes. Preliminary analyses confirmed that the speaker's occupation had no effect on participants' responses, hence this variable was not included in any of our main analyses.

each speaker and their relative veracity was counterbalanced across participants. This created a total of 16 lists in the experimental phase.

The 20 critical trials in each list were randomly presented together with 40 fillers, each also naming a referent as the object concealing the treasure. To increase variability, half of these included one of various forms of disfluency, or a discourse manipulation such as a non-propositional sentence marker or a modal varying the speaker's commitment to the truth value of the statement. The breakdown of filler utterance types is identical to that in Experiments 4.1 and 4.2 (see Table 4.1). Half the filler utterances were produced by one speaker and half by the other. As with the critical utterances, half of each speaker's fillers were fluent and the other half characterised by disfluency or some other form of manipulation.

In both the training and experimental phases, the two images on each visual display were centred vertically and positioned horizontally left and right on the screen. In each phase, referents were presented on either side of the screen half of the time. Distractor images used in each phase were drawn from a separate list of 60 objects and randomly paired with the referent, with no repetition of images across the experiment.

Procedure

The experiment was presented using OpenSesame (Mathôt et al., 2012) on a 21in. CRT monitor. Eye movements were monitored using an Eyelink 1000 Tower Mount system sampling at 500Hz. Mouse coordinates were sampled at 50Hz.

Figure 5.2 presents a sample of a trial from the experiment. Between trials, participants saw a central grey fixation dot and underwent a manual drift correct to ensure accurate recordings from the eyetracker. After this, the dot turned red for 500 ms to signify the start of the trial. Each trial began by presenting the identity of the upcoming speaker. The name of the speaker and their avatar appeared centred on screen for 1000 ms. This was replaced by the referent and

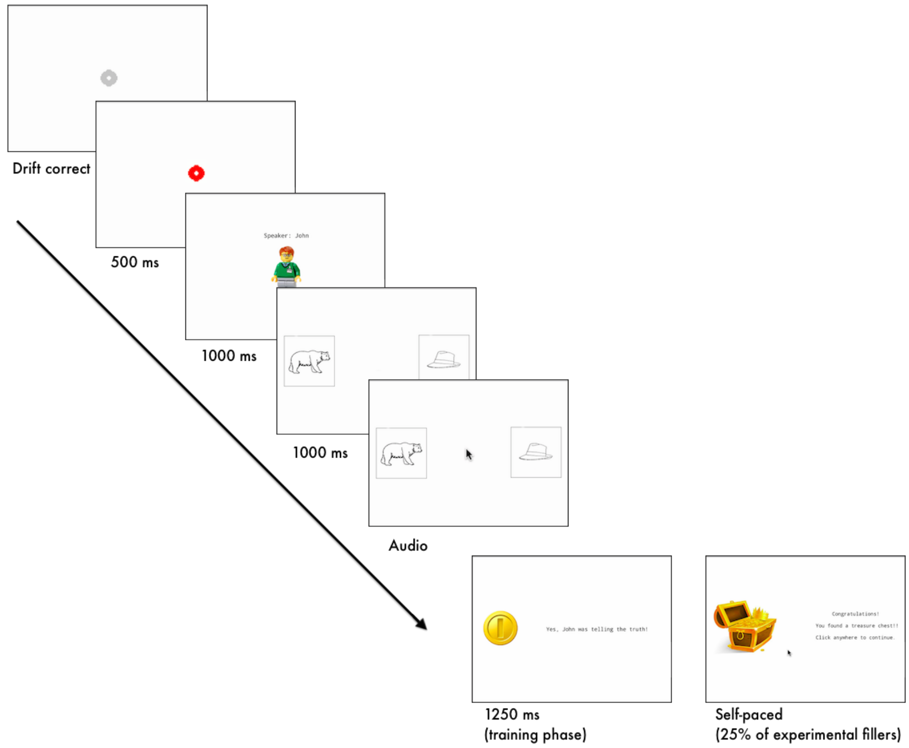


Figure 5.2: Experiment 5.1: Timeline of an example trial

distractor that made up each visual display. The display appeared for a 1000 ms preview, after which the mouse pointer appeared at the screen centre and playback of the audio stimulus began. Participants were instructed to click on the object that they believed concealed the treasure. The display disappeared once a mouse click had been recorded on one of the two objects.

On training trials, participants received feedback after each mouse click informing them of the trial outcome (treasure found or not found) and reinforcing the speaker's veracity on their previous utterance (e.g., *"Yes, Fred was lying!"* or *"Oh no, looks like John was telling the truth!"*). Ten points were added to the participant's score upon successful location of the treasure. Feedback remained on screen for 1250 ms, after which the grey fixation dot appeared to begin the next trial. As with the experiments in Chapter 4, feedback was not offered during the experimental phase, except on "bonus" filler trials (25%; see section 4.2.4). In both phases, trials had an automatic time-out 5 seconds post-utterance offset. If an object click was not detected before this, participants saw a message telling them to respond more quickly.

To reduce fatigue, the experiment included a break between the two phases. The experimental phase began when participants were ready to continue. The eye-tracker was then recalibrated, and participants were reminded that they would no longer receive feedback, except in the case of bonus trials. After the task, a post-test questionnaire was administered to assess whether participants had been sensitive to the speaker veracity manipulation in the training phase, and to verify that none had suspected that the audio stimuli had been scripted for purposes of the experiment. Participants were also orally questioned on these two points during debrief, before the actual purpose of the experiment was revealed.

5.2.2 Results

Statistical analyses were carried out in R version 3.3.3 (R Core Team, 2017) using the lme4 package (Bates et al., 2014). Separate analyses were conducted for the training and the experimental phase.

Analysis of the training phase data evaluated whether listeners successfully inferred that the two speakers differed in terms of veracity. To determine whether participants discriminated between the two speakers by the end of the phase, logistic mixed effects regression was used to model the binary outcome of final object click (referent/distractor), with speaker (truthful/dishonest) as the fixed effect and random intercepts and slopes for subjects and referents. To determine whether speaker discrimination unfolded over time, a model was fit using empirical logit regression (Barr, 2008), taking as the dependent variable the e-logit of the proportion of clicks recorded on the referent for each speaker on each trial. Speaker and trial were included as fixed effects, with random intercepts and slopes for both predictors by subjects and random intercepts by referents.

Analysis of the experimental phase data focused on participants' off-line (final object click) and on-line (eye and mouse movements) interpretations of speakers' utterances. Logistic mixed effects regression was used to model participants' final object click, with manner of delivery (fluent/disfluent) and speaker (truthful/dishonest) as fixed effects and by-subjects and by-referents random intercepts and slopes for both predictors.

Eye-tracking records were averaged into 20 ms bins, each comprising ten samples, prior to analysis. Fixation data were coded in terms of region of interest (referent/distractor/none), and the proportion of fixations to each object out of the total sum of fixations was computed for each time bin. Mouse-tracking analysis only took into account the X coordinates. For each sample, the distance travelled by the mouse was computed by taking the absolute difference between the X coordinates of the current and previous sample. The data were coded for direction of movement (toward referent/toward distractor) for each bin, and the cumulative distance participants had moved the mouse toward either object was calculated by summing over the distance travelled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). For each object, we then calculated a proportion-of-movement measure, defined as the distance travelled by the mouse pointer toward the given object, divided by the total distance travelled (regardless of X direction).

Model analyses for eye and mouse movements were conducted over a time window beginning from referent onset to 800 ms post-onset. Models were fit using empirical logit regression, taking as the dependent variable the difference between the e-logit of fixations (or mouse movements) to the referent and the distractor on each trial. Fixed effects included time, manner of delivery and speaker, with by-subjects and by-referents random intercepts and slopes for all predictors.

For both the training and experimental phase analyses, all predictors were mean-centred in all models.

Training phase

Table 5.1: Experiment 5.1: Breakdown of mouse clicks recorded on each object (referent or distractor) during utterances produced by each speaker (truthful or dishonest) in the training phase. Values represent percentage of trials.

	Referent	Distractor
Truthful speaker	66	34
Dishonest speaker	48	52

Table 5.1 presents the breakdown of mouse clicks recorded on each object for each speaker during the training phase. A logistic mixed effects regression on the outcome of mouse clicks showed an effect of speaker. Participants were more likely to click on the referent following utterances produced by the truthful speaker, $\beta = 0.77$, SE = 0.18, $p < .001$.

Fig. 5.3 shows the proportion of clicks recorded on the referent following utterances produced by each speaker over the course of the training phase. The pattern of responses over time suggests a difference by speaker, with a falling trend in the proportion of clicks to the referent for the dishonest speaker. This difference was reflected in a speaker by trial interaction, $\beta = -0.01$, SE = 0.004, $t = -2.309$, establishing that participants' likelihood of clicking on the referent following an utterance by the dishonest speaker decreased as the training phase progressed. These results confirm that participants did discriminate between the relative veracities of the two speakers as the training phase progressed.

Experimental phase: Final object click

Table 5.2 shows the breakdown of mouse clicks on each object following fluent and disfluent utterances produced by each speaker during the experimental phase. A logistic mixed effects regression revealed main effects of manner of delivery and

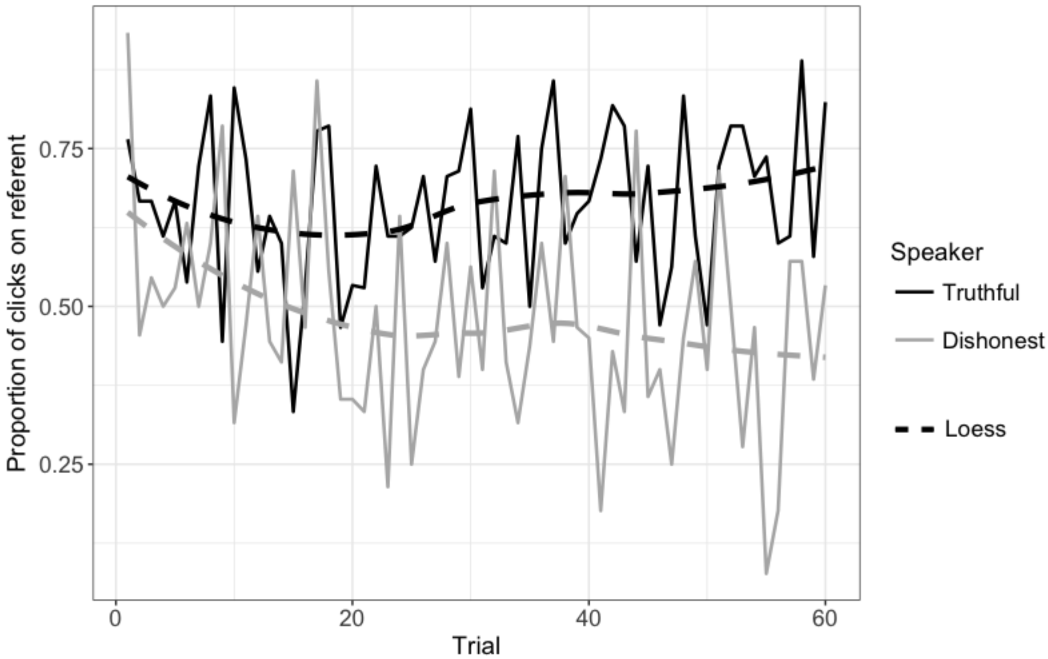


Figure 5.3: Experiment 5.1: Proportion of clicks to the referent for each speaker over trials during the training phase. The dashed lines represent the trial-by-trial data with a loess smoothing method applied.

speaker. Participants were more likely to click on the referent following a fluent utterance, $\beta = 1.55$, SE = 0.30, $p < .001$, as well as following an utterance by the truthful speaker, $\beta = 0.51$, SE = 0.23, $p = 0.03$. There was no interaction between speaker and manner of delivery to suggest a difference in participants' perception of fluent and disfluent utterances by each speaker ($p > .1$).

Experimental phase: Eye movements

Fig. 5.4 shows the proportion of fixations to each object following fluent and disfluent utterances by each speaker from noun onset until 2000 ms post-onset. The time course of fixations demonstrates a strong preference for participants to look at the referent following fluent utterances. This preference emerged rapidly, almost immediately post-onset of the noun and prior to completion of the utterance. In contrast, disfluent utterances were characterised by a decrease in looks

Table 5.2: Experiment 5.1: Breakdown of mouse clicks recorded on each object (referent or distractor) during fluent and disfluent utterances produced by each speaker in the experiment phase. Values represent percentage of trials.

		Referent	Distractor
Truthful speaker	Fluent	78	22
	Disfluent	42	58
Dishonest speaker	Fluent	64	36
	Disfluent	38	62

to the referent and an increase in looks to the distractor. Comparison of the eye movements reveals a generally similar pattern of fixations across the two speakers, with the decrease in looks to the referent following disfluent utterances slightly more pronounced for the truthful than the dishonest speaker. Analyses revealed a time by manner of delivery interaction, $\beta = 4.12$, SE = 0.23, $t = 17.26$, reflecting the increase in participants' fixation bias to the referent following fluent compared to disfluent utterances. There were no effects of speaker nor its interaction with any of the other variables, suggesting that the effect of manner of delivery did not differ for utterances produced by the truthful and the dishonest speaker.

Experimental phase: Mouse movements

Fig. 5.5 shows the proportion of mouse movements (in terms of distance travelled) toward each object during fluent and disfluent utterances by each speaker from noun onset until 2000 ms post onset. The time course of mouse movements indicates a bias by manner of delivery compatible with the fixation data. During fluent utterances, participants were overwhelmingly more likely to move the mouse toward the referent than the distractor for both speakers. This difference emerged rapidly, around 400–500 ms post-onset of the noun in the utterance. During disfluent utterances, participants' mouse movements displayed a slight

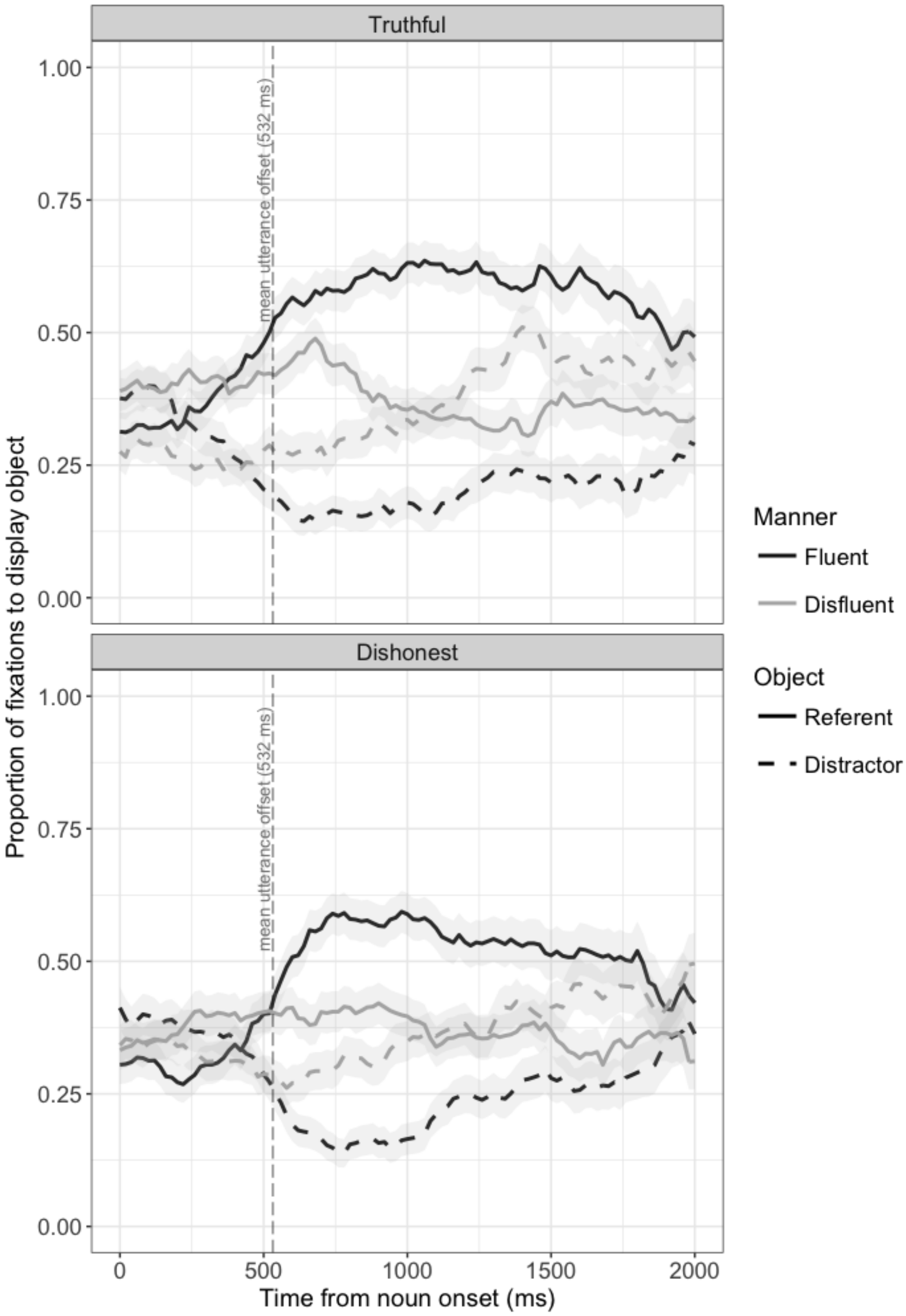


Figure 5.4: Eye-tracking results for Experiment 5.1: Proportion of fixations to each object (referent/distractor) over time for fluent/disfluent utterances for each speaker. Shaded areas represent ±1 standard error of the mean.

preference toward the distractor that emerged later, with this preference qualitatively more pronounced following utterances by the truthful speaker. The difference in mouse movements following fluent and disfluent utterances was confirmed by a time by manner interaction, $\beta = 5.29$, SE = 0.37, $t = 14.18$. As with the eye movements, there were no effects of speaker nor its interaction with any other variables to suggest a difference in the effect of manner across the two speakers.

Experimental phase: Post-hoc analyses

To rule out the possibility that the results were due to participants dissociating the second phase of the experiment from the first (thereby effectively "resetting" any expectations based on relative speaker veracity), a subsidiary analysis was conducted to compare the subset of critical trials encountered during the first third of the experimental phase and those encountered during the last third. Participants' final object clicks showed a greater referent-bias during utterances produced by the truthful speaker during the first third trial subset compared to the last third, as evidenced by a marginal speaker by trial subset interaction, $\beta = -0.80$, SE = 0.43, $p = 0.07$. Participants' eye movements also exhibited a different pattern by trial subset. Participants were less likely to fixate the referent over the distractor following utterances by the truthful speaker in the last third compared to the first third, as evidenced by a time by speaker by subset interaction $\beta = -2.22$, SE = 0.59, $t = -3.79$. Moreover, participants were more likely to fixate the referent over the distractor following fluent utterances in the last third compared to the first third, as evidenced by a time by manner by subset interaction $\beta = 1.38$, SE = 0.17, $t = 8.12$. Taken together, this pattern suggests that participants did indeed carry over the speaker bias established in the training phase, but this was superseded by a disfluency bias over the course of the experimental phase.

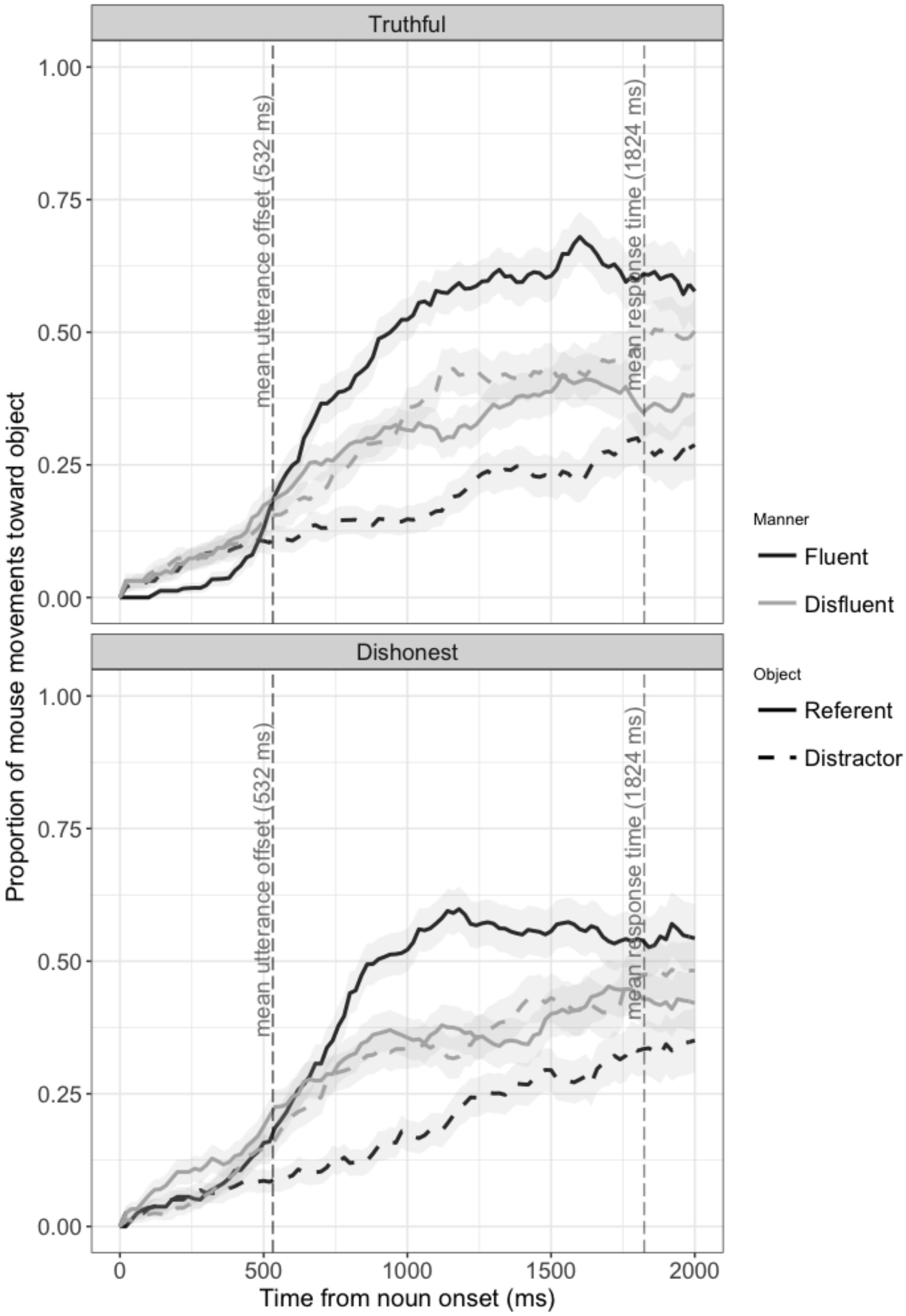


Figure 5.5: Mouse-tracking results for Experiment 5.1: Proportion of cumulative distance travelled toward each object (referent/distractor) over time for fluent/disfluent utterances for each speaker. Shaded areas represent ± 1 standard error of the mean.

5.2.3 Discussion

Our results establish that listeners make pragmatic inferences about a speaker's veracity based on their manner of delivery, with disfluency biasing listeners toward interpreting an utterance as a lie. These results corroborate findings from Experiments 4.1 and 4.2. The time course of eye and mouse movements similarly reveals that this bias unfolded rapidly, within hundreds of milliseconds from the point in the utterance that listeners could infer whether the speaker had lied.

Results from the training phase suggest that listeners developed expectations based on each speaker's veracity over time. This was further confirmed by results from the post-test questionnaire and conversations at debrief, which revealed that 21 out of 32 participants indicated an explicit awareness of how likely each speaker was to lie.[2] This provides surface evidence that our speaker manipulation was effective, such that listeners discriminated between the two speakers in terms of perceived veracity. More broadly, these results also align with findings from statistical learning that individuals can grasp complex information over time through exposure to distributional patterns. For example, learners rapidly acquire grammatical rules (Romberg & Saffran, 2010; Wonnacott, Newport, & Tanenhaus, 2008) or word meanings (Mirman, Magnuson, Estes, & Dixon, 2008; Vouloumanos, 2008) based on probabilistic distributions in the data, even under learning conditions where input may be highly ambiguous (Yu & Smith, 2007). Here, we demonstrate that such systematic acquisition of information can extend to metalinguistic knowledge such as properties relating to a specific speaker (cf. Yu & Ballard, 2007).

The referent-bias observed in the object clicks for the truthful compared to the dishonest speaker during the experimental phase suggests that listeners carried over their expectations about relative speaker veracity from the training phase.

[2]Note, additionally, that comprehenders do not always express overt awareness when they pick up on differences between speakers (cf. Grodner & Sedivy, 2011).

This was confirmed by the effect of speaker on listeners' final object clicks in the first third of experimental trials, which demonstrates that listeners were more likely to expect the dishonest speaker to lie. Given that, it is perhaps surprising that these expectations did not affect how listeners perceived each speaker's disfluencies.

There are a number of potential explanations for this finding. The first and simplest is that listeners simply lost their speaker biases once they stopped receiving feedback during the experimental phase. This is unlikely given the results from our post-hoc analyses to suggest speaker discrimination persisted into the first third of the experimental phase. A second possibility relates to the nature of the cover story, which established a context in which utterances, both true and false, were produced with the overall aim of misleading a listener into clicking on the wrong location. Consequently, within this (fictitious) context, truthful utterances could have been produced by speakers with deceptive intent. It is possible that under a cognitive hypothesis to deception, listeners may have perceived truth-telling (with deceptive intent, i.e., double-bluffing) as more effortful than straightforward lying. Listeners might additionally have reasoned that this would be harder for a truthful speaker than a dishonest one, creating a bias to interpret the truthful speaker's disfluencies as an index of truth-telling. This may in turn have resulted in similar rates of disfluency-lie judgements across the two speakers, albeit via distinct reasoning mechanisms. It is difficult, based on the current data, to provide evidence for or against this account. However, in Chapter 6 we discuss this possibility in greater detail in the context of an experiment that provides immediate feedback on listeners' veracity judgements based on a speaker's manner of delivery.

The third, and most likely possibility given our results, is that the disfluency-lying bias overruled the speaker-specific expectations acquired during the training phase. In other words, such associations based on a speaker's manner of delivery may be so ingrained in listeners (Akehurst et al., 1996; Zuckerman, Koestner, &

Driver, 1981), that they dominate even in the face of other sources of information that can indicate deceptive intent. This explanation is consistent with evidence from our post-hoc analyses to suggest a stronger effect of speaker in the first third compared to the last third of the experimental phase, in both listeners' eye movements and object clicks. In contrast, listeners' eye movements indicate a stronger effect of speakers' manner of delivery in the last third compared to the first third. Together, this pattern of results suggests that listeners' speaker-dependent inferences about utterance veracity were superseded by a stronger bias to perceive a lie based on disfluency over the course of the experimental phase. In the next experiment, we turn our attention to a different form of speaker-specific expectations: that relating to a speaker's accent.

5.3 Experiment 5.2

Few studies to date have examined the effect of a speaker's accent on a listener's perception of deception. What little research that exists has focused on the ability in listeners to accurately detect deception in non-native speakers. These studies suggest a higher accuracy at judging deception in one's native language, as well as a lie-bias when evaluating non-native speech (Da Silva & Leach, 2013; Evans & Michael, 2014). This may be due to a tendency to perceive non-native speakers in a negative light (e.g., Rubin & Smith, 1990), but could as well reflect a more general processing disadvantage for non-native speech (Munro & Derwing, 1995), affecting how listeners rely on typical diagnostic cues such as a speaker's manner of delivery. For example, research on the properties of non-native speech highlight its susceptibility to disfluency (Cucchiarini, Strik, & Boves, 2002; Tavakoli, 2010), due to various increased demands on the production system (Segalowitz, 2010). This might reduce or eliminate disfluency's cue potential as an indicator of lie behaviour, as it is construed as an index of a range of causes linked to production difficulty. On the other hand, under an account that emphasises the effort associated with deception (Vrij & Granhag, 2012), non-native speakers might be expected to be less capable of coping with the additional demand of lying on top of speaking a foreign language, increasing the likelihood that disfluency indexes lying.

More generally, research outside of deception suggests that listeners adapt their disfluency-dependent biases based on their expectations relating to a speaker's accent (Bosker et al., 2014). However, it remains unclear how, in the context of a pragmatic inference such as judging deception, listeners' perception of a speaker's disfluencies would be influenced by the speaker's accent.

The current experiment was designed to test whether listeners' disfluency-lying bias is modulated by their expectations based on a speaker's accent. The paradigm used in Experiments 4.1 and 4.2 was modified to include two speakers, a native and a non-native speaker of English. As before, listeners' aim on each trial was to locate hidden treasure based on the speaker's description, under the guise that the descriptions were obtained from an earlier experiment in which speakers were incentivised to lie some of the time about the treasure's location.

5.3.1 Method

Participants

Thirty-two self-reported native British English speakers took part in the experiment, all of whom fulfilled the same requirements as Experiment 5.1. None had participated in any of the earlier experiments. All participants provided informed consent in accordance with the university's Psychology Research Ethics Committee guidelines (Ref. No.: 6-1617/1).

Materials and design

The visual materials and design were identical to the experimental phase of Experiment 5.1. Audio stimuli were replaced by a new set of recordings produced by two speakers (both female)—a native (British) and a non-native (L1 Italian) speaker of English. Each voice as associated with a Lego avatar (see Fig. 5.6), counterbalanced across participants such that the avatar associated with native accent in half the participants was associated with the non-native accent for the other half.

Figure 5.6: Experiment 5.2: Lego avatars used to represent the two speakers. The accent associated with each avatar was counterbalanced across participants.

The speakers were recorded individually in a sound-attenuated booth. For the non-native speaker, the recording procedure of Bosker et al. (2014) was adopted: The speaker listened to a playback of the native speaker's speech, which she then imitated, utterance by utterance. This was done to minimise all contrasts other than accent between the two speakers as much as possible. For both speakers, utterances were initially recorded in their entirety; the disfluency segment from each disfluent utterance was then cross-spliced into its corresponding fluent utterance to create a disfluent counterpart. Utterances were normalised to have the same mean acoustic intensity. Table 5.3 provides a summary of the prosodic properties of each speaker's utterances.

The 20 critical items were counterbalanced across four lists in a 2 (fluent/disfluent delivery) x 2 (native/non-native accent) design. These were presented alongside 40 fillers, characterised by the same breakdown as in Experiment 5.1.

Procedure

The experiment procedure was identical to the experimental phase of Experiment 5.1. After the task, a questionnaire was administered to verify that participants

Table 5.3: Experiment 5.2: Duration (ms) and pitch (Hz) properties of fluent and disfluent utterances by each speaker

		Mean utterance dur.	Mean referent dur.	Mean pitch	Max. pitch
Native	Fluent	1688	578	221	278
	Disfluent	2747	578	206	299
Non-native	Fluent	1641	604	203	275
	Disfluent	2702	604	191	287

had not suspected the cover story or that the audio had been scripted. Participants were also orally questioned on this point during debrief, before the actual purpose of the experiment was revealed.

5.3.2 Results

We followed the same analysis procedures that we used on the experimental phase data in Experiment 5.1. Data from 3 experimental trials (0.5%) on which a click was not recorded on either object were excluded from the final dataset.

Final object click

Table 5.4 shows the breakdown of mouse clicks on each object during fluent and disfluent utterances produced by each speaker. A logistic mixed effects regression revealed main effects of manner of delivery and accent. Participants were more likely to click on the referent following a fluent utterance $\beta = 1.97$, SE = 0.45, $p < .001$, and less likely to click on the referent following an utterance produced by the native speaker $\beta = -0.58$, SE = 0.19, $p < .01$. There was no interaction between the two factors to suggest that participants' disfluency-based inferences about utterance veracity were affected by the speaker's accent.

Table 5.4: Experiment 5.2: Breakdown of mouse clicks recorded on each object (referent or distractor) during fluent and disfluent utterances produced by each speaker in the experiment phase. Values represent percentage of trials.

		Referent	Distractor
Native speaker	Fluent	67	33
	Disfluent	28	71
Non-native speaker	Fluent	74	26
	Disfluent	41	58

Note that percentages in some rows do not sum to 100 since clicks were not recorded on 0.5% of trials.

Eye movements

Fig. 5.7 shows the proportion of fixations to each object following fluent and disfluent utterances by each speaker from noun onset until 2000 ms post-onset. The time course of fixations reveals a difference based on the speaker's manner of delivery. During fluent utterances, participants demonstrate a fixation bias to the referent over the distractor that develops rapidly post-noun onset. Conversely, during disfluent utterances, after a brief initial rise in looks to the referent, this decreases and a corresponding rise in looks to the distractor emerges. A comparison across speakers also suggests a difference by accent, with a greater overall preference to fixate the referent over the distractor during non-native compared to native speech. These differences are reflected in our model analyses, which reveal a time by manner interaction $\beta = 3.23$, SE $= 0.24$, $t = 13.27$ as well as a time by speaker interaction $\beta = -0.71$, SE $= 0.24$, $t = -2.93$. There was no evidence of a time by manner by speaker interaction to suggest that the effect of manner of delivery differed between native and non-native speech.

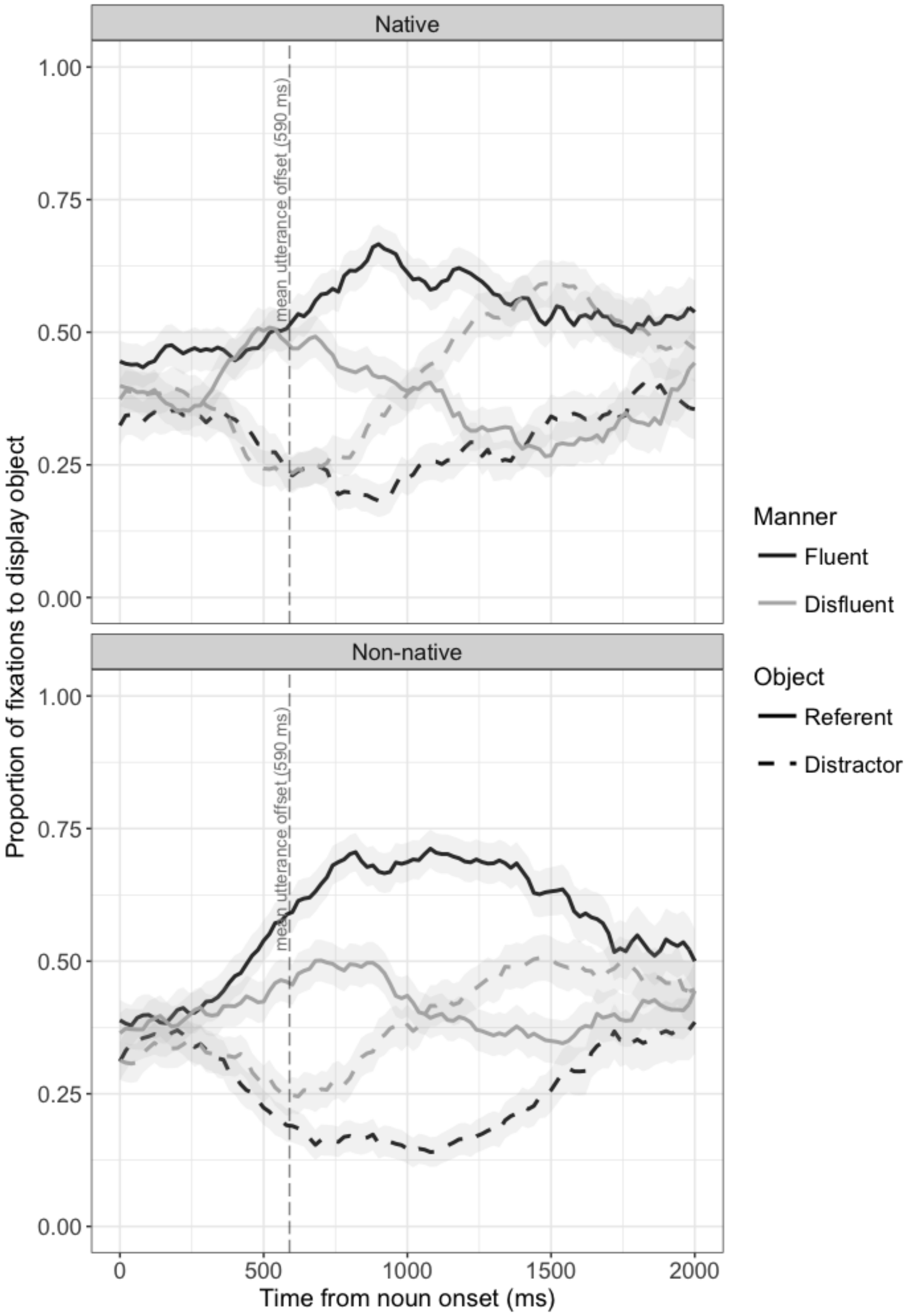


Figure 5.7: Eye-tracking results for Experiment 5.2: Proportion of fixations to each object (referent/distractor) over time for fluent/disfluent utterances for each speaker. Shaded areas represent ±1 standard error of the mean.

Mouse movements

Fig. 5.8 shows the proportion of mouse movements toward each object following fluent and disfluent utterances by each speaker from noun onset until 2000 ms post-onset. As with the eye movements, participants' mouse movements suggest a difference by manner of delivery as well as by speaker. During fluent utterances, participants showed a strong preference to move the mouse toward the referent over the distractor, beginning around 500 ms after onset of the noun. In contrast, disfluent utterances reflect a smaller but distinct bias toward the distractor. This bias was more pronounced following utterances produced by the native speaker during the later part of the trial. Non-native utterances were also characterised by a greater overall referent-bias compared to native utterances. These differences were reflected in our model analyses, which revealed a time by manner interaction, $\beta = 10.20$, SE $= 0.41$, $t = 24.60$, as well as a time by speaker interaction $\beta = -1.94$, SE $= 0.41$, $t = -4.68$. This pattern of mouse movements is also consistent with participants' final object clicks, which indicate a preference to click on the referent following fluent compared to disfluent utterances, and non-native compared to native speech. There was no evidence of a time by manner by speaker interaction to suggest that the effect of manner of delivery was modulated by the speaker's accent.

Post-hoc analyses

Based on the observation that participants' later mouse movements exhibited a stronger distractor bias following native disfluent utterances (relative to non-native), we conducted a secondary analysis to determine whether the effect of manner differed between the two speakers during the later part of the trial. Analyses over a 800–1600 ms time window revealed a significant time by manner by speaker interaction, $\beta = -4.84$, SE $= 1.28$, $t = -3.79$, reflecting a weaker disfluency-lying bias for non-native speech during the later stages of comprehen-

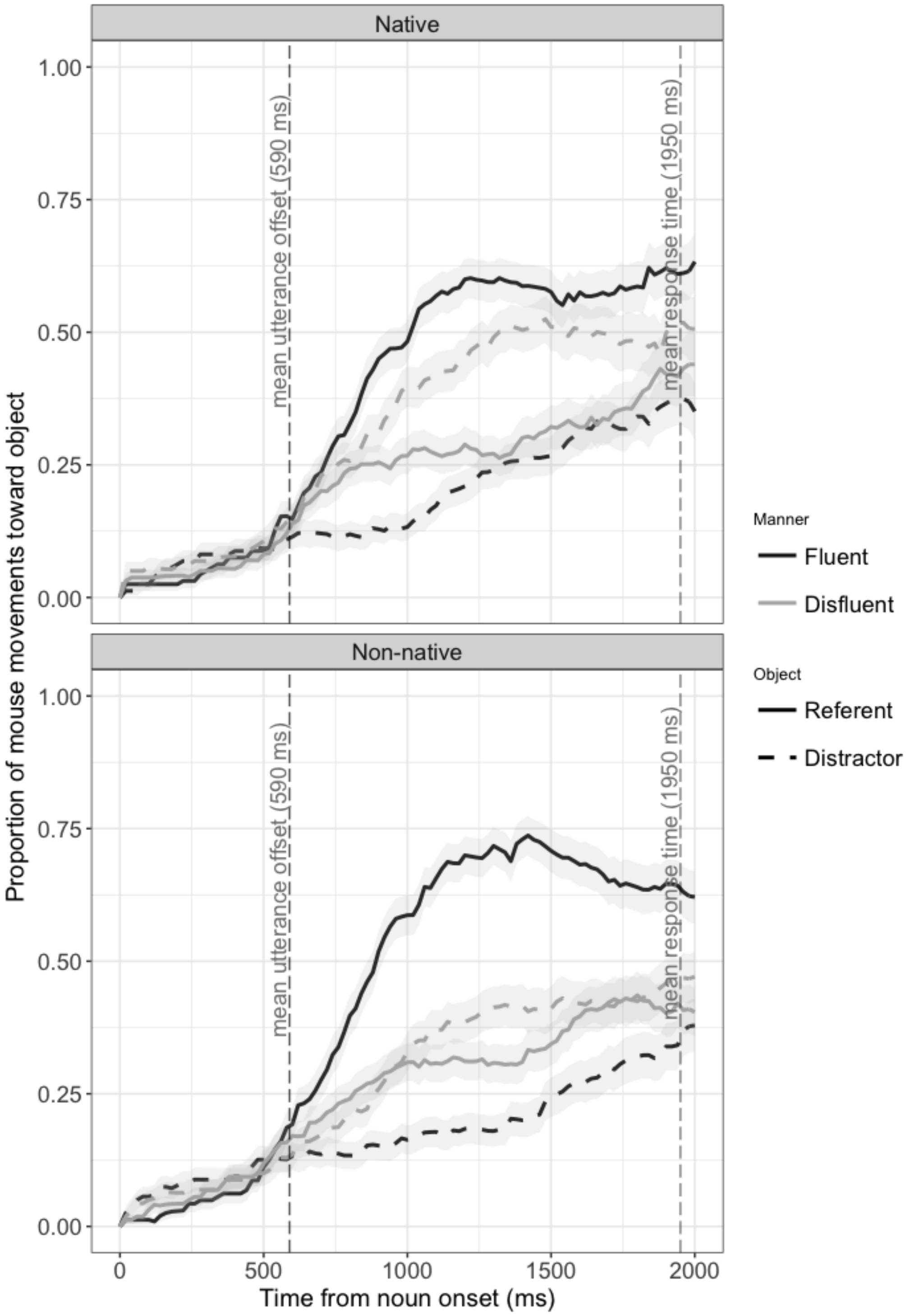


Figure 5.8: Mouse-tracking results for Experiment 5.2: Proportion of cumulative distance toward each object (referent/distractor) over time for fluent/disfluent utterances for each speaker. Shaded areas represent ±1 standard error of the mean.

sion. Intriguingly, this effect was not present in participants' eye movements over the same time window, $\beta = -0.28$, SE = 0.50, $t = -0.55$. We return to this result in the Discussion.

5.3.3 Discussion

In line with our earlier experiments, results from Experiment 5.2 show that listeners' pragmatic inferences about utterance veracity were rapidly influenced by the speaker's manner of delivery. Fluent utterances led to a truth-bias, while disfluent utterances revealed an opposite lie-bias. This bias held regardless of whether the utterance was produced by a native or a non-native speaker of English, and highlights listeners' rapid sensitivity to a speaker's manner of delivery in their on-line pragmatic hypotheses about the speaker's veracity.

Listeners' veracity inferences were additionally influenced by the speaker's accent. This is broadly consistent with studies on accent perception, which suggest that listeners hold pre-existing attitudes toward non-native accents, which subsequently affect their perception of what the speaker says (Gluszek & Dovidio, 2010; Rubin & Smith, 1990). Surprisingly, our accent effects were in the opposite direction to those observed by earlier studies on deception. These studies report a bias in native listeners to believe non-native speakers are lying (Da Silva & Leach, 2013; Evans & Michael, 2014); in contrast, both our on-line and off-line data suggest that listeners were more likely to interpret non-native utterances as truthful compared to native ones. These conflicting results are difficult to reconcile, but we note a diversity in the non-native accents used by lie perception studies thus far. This might implicate a role for language or culture-specific stereotypes in the perception of non-native deception (cf. C. F. Bond et al., 1990). Another explanation could be differences in the type of lie at stake. One reason for a lie prejudice against non-native speech is the emotional detachment associated with speaking in a second language (M. H. Bond & Lai, 1986; Dewaele, 2004). In particular, lying in one's first language evokes a stronger emotional response than lying in a second language (Caldwell-Harris & Ayçiçeği-Dinn, 2009). From a listener's perspective, this may elicit a lie-bias against non-native speakers, particularly in the context of lies which have an element of self-connection, such as the ones used in Evans and Michael (2014). However, under this account, it is unlikely for listeners to hold a similar bias for utterances that lack emotional content, such as the ones used in our experiment. More generally, the inconsistency across studies calls for more in-depth research on lie perception in non-native speech, taking into account factors that may be specific to the language or the discourse.

Finally, we found little evidence that listeners' expectations based on the speaker's accent influenced their tendency to attribute disfluency to lie-telling. In both the off-line responses and the early eye and mouse movements, a strong disfluency-lying bias was observed across native and non-native utterances. However, this bias appeared to differ by accent during the later part of the trial, with participants' mouse movements demonstrating a stronger lie bias following disfluent utterances by the native speaker. Analysis of listeners' eye movements, however, did not indicate a comparable bias over the same time window. We do not currently have an explanation for this discrepancy, although we suggest that it might reflect general differences between the nature of eye and mouse movements. The saccadic nature of eye movements means that they tend to exhibit discrete "jumps" between objects; in contrast, mouse movements typically follow a fluid, continuous motion across space (see section 2.4.2). The absence of any modulating effect of accent in listeners' final object clicks suggests that the difference by accent was transient, and ultimately overridden by a stronger bias to associate disfluency with lie-telling regardless of the speaker.

5.4 Chapter discussion

The two experiments reported in this chapter set out to test whether listeners' manner-based inferences about speaker veracity are modulated by their expectations about the particular speaker. Overall, we found no evidence for a speaker-specific effect of the disfluency-lying bias. This finding is at odds with the literature on prediction, which provides strong evidence for speaker-specificity in the disfluency bias observed during referential prediction (J. E. Arnold et al., 2007; Barr & Seyfeddinipur, 2010; Bosker et al., 2014). This inconsistency may point to differences in the cognitive processes underlying disfluency comprehension across different contexts. The effect in prediction, for example, has been attributed to a perspective-taking mechanism, by which listeners form associations for an *um* based on their perceived cognitive state of a given speaker (Barr & Seyfeddinipur, 2010). It is possible, however, that such perspective-taking mechanisms about the source of a disfluency are harder for listeners to extend to their pragmatic inferential processes, such as judgements in contexts where a speaker's veracity may be called into question.

Although we found no evidence for a speaker-specific effect of disfluency, It remains notable that the disfluency-lie association was the prevailing strategy adopted by listeners across both experiments. This effect persisted despite the fact that listeners received no indication to confirm or disprove the bias, and moreover held in the presence of other sources of information (expectations about a speaker) that could influence their bias. This is particularly notable in Experiment 5.1, where feedback during the training phase meant that listeners formed clear hypotheses about each speaker's likelihood of lying. The fact that listeners later attended to speakers' disfluencies to infer lie behaviour, rather than drawing on prior knowledge about the speaker's veracity, points to the dominant nature of these verbalisations as a cue to deception. This is further evidenced by the speed with which listeners' inferences unfolded, with disfluency driving interpretation almost as soon as the speaker's intended meaning could be inferred. These results are counter to predictions by the standard pragmatic model, which maintains that pragmatic inferences occur at a later, post-literal stage of processing. Instead, they delineate the immediacy with which information based on the speaker's manner of delivery was extracted and used to constrain listeners' judgements of veracity. Extending our results from Chapter 4, we show that these rapid, manner-based inferences remain robust, even in the face of alternative information that could potentially guide a listener's response.

Overall, our findings align with with broader evidence that listeners rapidly reason about a speaker's manner of delivery, which in turn affects how they interpret the unfolding speech. The outcome of this reasoning points toward a cognitive account by which disfluency is attributed to speaker difficulty, whether the result of describing difficult objects (e.g., J. E. Arnold et al., 2007), or from having to formulate a lie (current experiments). The degree of flexibility of this process (e.g., whether or not listeners are able to adapt this bias to their expectations about the given speaker) may depend on the context of occurrence, as well as other concurrent sources of information. In the context of perceiving deception, for example, the bias to infer a lie based on disfluency may be so ingrained in listeners, that it may hold universally across speakers, and regardless of whether it is substantiated by immediate evidence.

This leads to the question, then, of whether such biases in fact have a basis in reality. Are they valid associations, founded on actual observation of liars in the real world? Or are they in fact the product of (misguided) reasoning, arising perhaps from folk notions of deception (cf. Heydon, 2008)? In the next chapter, we address this question by examining the disfluency-lie bias from both a production and a perception perspective. By adapting the current paradigm into an interactive game of lying and lie detecting, we explore whether the stereotypes of deceit listeners hold are in line with the actual lie behaviour that speakers exhibit.

Chapter 6

Manner of delivery in real-life deception[1]

In the previous two chapters, we established that listeners rapidly infer that a speaker is lying when the speaker is disfluent in their manner of delivery. Moreover, we found that these manner-based inferences hold regardless of the locus of the disfluency, and are robust even in the face of other sources of information which may modulate the effect. These findings are largely consistent with the earlier literature on deception, which frequently highlights disfluencies as an index of lie behaviour (Sporer & Schwandt, 2006; Zuckerman, Koestner, & Driver, 1981). Taken together, these results point to the strength of the association, to the extent that it may be difficult for listeners to overcome, even when presented with alternative, potentially more reliable, evidence.

In this chapter, we test the robustness of the disfluency-lie bias in an ecological context where participants interact in real-time conversation. We adapt the treasure game comprehension paradigm into an interactive, dyadic game of

[1]This chapter constitutes an extended version of a manuscript that was recently submitted to *Cognitive Science* (Loy, J. E., Rohde, H. & Corley, M. Cues to lying may be deceptive: Speaker and listener behavior in an interactive game of deception.) The chapter includes an extended analysis and discussion of listeners' mouse movement data. This was recorded during the experiment but not included in the manuscript as it was outwith the scope of the study. The findings are discussed here as they are of interest in relation to the questions concerning the time course of pragmatic inferencing that this thesis addresses.

deception, in which speakers produce a mix of true and false utterances, and listeners judge the veracity of each utterance as it unfolds. This allows us to study speaker and listener behaviour concurrently during the real-time production and perception of deception. The motivation of the study was two-fold: To test whether a) listeners' associations between disfluency and lying still hold in a more naturalistic setting, and b) these manner-based associations align with speakers' actual behaviour during lie production. The structure of the chapter is as follows. We first outline the motivations of the study in greater detail, with reference to the relevant literature on deception and communication. We then present an exploratory study investigating the role of manner of delivery in speakers' production and listeners' perception of deception. This is followed by an analysis and discussion of listeners' mouse movements, which were measured during the experiment but not reported in the main study. Finally, we discuss our findings from both the exploratory study and the mouse-tracking analyses in relation to the broader questions of this thesis.

6.1 Investigating the production and perception of deception

Thus far, our finding that disfluency rapidly biases listeners toward inferring a lie has been robust across studies, reflecting the deep-rooted nature of this association. The experiments so far, though, have limited their investigations to noninteractive contexts, focusing solely on the listener's perspective. Deception is, however, a largely social phenomenon: As Burgoon, Buller, Floyd, and Grandpre (1996, p.726) point out, "deception often transpires in interactive circumstances, where both senders and receivers are active participants ... and where strategies by which senders maintain their deception and receivers' or observers' attempts to detect deceit can change over the course of an interaction". Overlooking the communicative element consequently results in a one-sided view of the event, under-representing what could take place in a real-life deception scenario.

Evidence that listeners may behave differently in interactive situations comes from studies which highlight differences in the quality of understanding for interlocutors compared to third-party overhearers (e.g., Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). This suggests that the act of conversational involvement may play a role in influencing the process of comprehension. One reason for this difference could be the availability of additional sources of information in interactive contexts, such as cues in both the visual and auditory channels, forefronted by the immediacy of conversational demands. In the context of deception, this becomes particularly relevant when considering the spectrum of cues that listeners can potentially draw on (DePaulo et al., 2003), raising questions about which cues hold greatest perceptual relevance in the face of multiple competing cues. Some studies suggest, for example, that speech cues hold greater discriminative value than nonverbal behavioural cues (e.g., DePaulo et al., 1982), while others highlight the relevance of attending to speakers' nonverbal behaviour (e.g., Ekman & Friesen, 1974). More generally, listeners' judgements of deception vary depending on whether they have access to both visual and auditory information, or only information in one modality (C. F. Bond & DePaulo, 2006; cf. Swerts & Krahmer, 2005). With these considerations in mind, we explore listeners' perception of cues to lying in a multi-modal, interactive situation, using a dialogic task that simulates a naturalistic, conversational setting. In addition to testing whether the disfluency-lie association still holds, we examine a range of cues based on speakers' manner of delivery to determine which cues listeners attend to when judging utterance veracity.

As mentioned previously, the robustness of listeners' disfluency-lie bias also calls into question whether these associations are reflected in the actual lie behaviour of speakers. In particular, the fact that these biases prevail despite the absence of any external validation suggests that these pragmatic inferences arise via some process of internal reasoning about human social cognition (cf. Goodman & Stuhlmüller, 2013). This reasoning appears to be built on a model of the speaker's production system where, under a cognitive account of deception (Vrij et al., 2008), lying is associated with cognitive effort, and cognitive effort results in disfluency (King et al., 2017). However, it is possible that this model is founded on listeners' beliefs concerning their own experience as a speaker (cf. Zuckerman, Koestner, & Driver, 1981), rather than their observation of other speakers around them. Do speakers in fact exhibit behaviour consistent with these biases that listeners hold? Moreover, in a context where listeners are presented with immediate feedback on the validity of their judgement biases, are they able to take this into account to adapt their biases in response to speakers' behaviour? In the final study of this thesis, we aim to address these questions by exploring deception during real-time interaction. Below, we present an experiment that situates the comprehension paradigm from earlier chapters within an ecological, interactive context. This allows us to investigate speakers' production and listeners' perception of manner-based cues to lying, and how these may be influenced by the conversational demands of interaction.

6.2 Experiment 6.1

6.2.1 Abstract

Are the cues that speakers produce when lying the same cues that listeners attend to when attempting to detect deceit? We used a two-person interactive game to explore the production and perception of speech and nonverbal cues to lying. In each game turn, participants viewed pairs of images, with the location of some treasure indicated to the speaker but not to the listener. The speaker described the location of the treasure, with the objective of misleading the listener about its true location; the listener attempted to locate the treasure, based on their judgment of the speaker's veracity. Our results demonstrate a surprising mis-

match between the cues that listeners perceive to index falsehood, and those that speakers produce when lying. Such a pattern is in keeping with the attempted control hypothesis, whereby liars may take into account listeners' expectations and correspondingly manipulate their behavior to avoid detection. Listeners' responses, however, suggest that they attend primarily to behaviors associated with increased mental difficulty, perhaps because lying, under a cognitive hypothesis, is thought to yield an increased cognitive load.

6.2.2 Introduction

In producing an utterance that is false, a speaker's behavior may contain cues which could signal their intention to deceive. These may range from speech cues such as hesitations, speech disturbances, and changes in the pitch or rate of speech, to nonverbal cues such as blinking and hand gestures (DePaulo et al., 2003; Sporer & Schwandt, 2006). Likewise, when evaluating the veracity of an utterance, a listener can draw on multiple cues to attempt to distinguish between truth and lie (Akehurst et al., 1996; Zuckerman, Koestner, & Driver, 1981).

The present study explores a range of behavioral cues to lying, with the aim of investigating whether the cues used by listeners to detect falsehood are the same as the ones that speakers produce when lying. We consider both speech and nonverbal cues since face-to-face communication is inherently multimodal, with information in both modalities contributing to shape the overall interpretation of an utterance. Speech cues such as intonation or disfluencies have been shown to influence a listener's interpretation of a speaker's metacognitive state (Krahmer & Swerts, 2005), while nonverbal cues such as facial expressions or gestural acts can influence perception of a speaker's emotional state (Busso & Narayanan, 2007) and facilitate pragmatic comprehension of the speech (Kirk, Pine, & Ryder, 2011). For example, listeners asked to estimate a speaker's certainty in answering a question achieve higher accuracy when presented with both speech and nonverbal information than when presented with information in only one modality, highlighting the perceptual relevance of multiple modes of information for listeners in deriving the overall meaning of an utterance (Swerts & Krahmer, 2005). This is particularly the case for the perception of lying, where a listener's final interpretation of the message necessarily goes beyond the surface meaning of the words. In the next section, we first discuss prior work examining speech cues to lying, looking at both the perception and production of cues, followed by a similar discussion on nonverbal cues. We compare models of deception that provide competing explanations regarding the presence of behavioral cues.

Behavioral cues to lying

A review of studies on the perception of speech cues to lying delineates a stereotypical image of a liar. Listeners expect liars to speak slower, pause longer, and speak with a higher pitched voice (Vrij, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). These cues appear to carry perceptual relevance whether subjects are evaluating lie behavior in themselves or in others (Zuckerman, Koestner, & Driver, 1981), reinforcing common depictions of one engaged in deceit. Chief among the set of belief cues that listeners associate with lying are filled pauses, or verbalizations such as *um* and *uh* that mark hesitation on the speaker's part. This interpretation follows the belief that *um* and *uh* arise from production problems (Clark & Fox Tree, 2002; Fox Tree, 2007), a trait commonly associated with lying (Vrij et al., 2001).

Do listeners' beliefs about cues to lying align with the actual cues produced by liars? From a production perspective, there is less clarity regarding the speech correlates of lying. Although most studies report a discriminative value in some cue behaviors, a comparison across studies yields conflicting results. For example, Vrij et al. (2001) observe an increase in the speech rate of liars compared to truth-tellers, while Strömwall, Hartwig, and Granhag (2006) report no reliable difference between the two groups. Both studies made use of a mock interrogation paradigm, in which participants were interviewed about a (staged) crime with instructions either to lie or to tell the truth, and both employed the same measure of rate of speech (number of words divided by the duration of time spent speaking). To add to the uncertainty, some meta-analytic reviews of lie production studies note a decrease (albeit non-significant) in the speech rates of liars compared to truth-tellers (e.g., Zuckerman, DePaulo, & Rosenthal, 1981).

Production studies also frequently identify filled pauses as a behavioral correlate of lying. However, studies disagree as to the direction of correlation. In line with listeners' expectations, several studies report an increase in filled pause production in liars (Vrij et al., 2001; Vrij & Winkel, 1991; Zuckerman, Koestner, & Driver, 1981). However, other studies report a decrease in liars' filled pauses (Arciuli et al., 2010; Vrij, 1995), and yet others report no reliable difference between liars and truth-tellers (Granhag & Strömwall, 2002; Mann, Vrij, & Bull, 2002). Such inconsistencies across studies highlight the challenges faced in establishing a coherent pattern, both in terms of the cues associated with lying and the direction of the relationship.

As with speech cues, previous studies identify a range of nonverbal indicators which people may employ to detect a liar. Lie perception experiments show that gaze aversion and fidgeting are behaviors commonly used to index falsehood (Mann et al., 2004; Vrij & Mann, 2001). These findings are largely consistent with results from questionnaires investigating peoples' beliefs about lying, which reveal that observers tend to interpret cues such as decreased eye contact and a higher frequency of adaptors (e.g., scratching, touching one's hair or clothing, and other self-directed manipulations) as signs of dishonesty (Akehurst et al., 1996; Vrij & Semin, 1996; Zuckerman, Koestner, & Driver, 1981). One possible reason why gaze behavior tends to correlate strongly with lie perception is due to the assumption that liars experience nervousness (Ekman, 2001; Knapp, Hart, & Dennis, 1974), and consequently avert their gaze from listeners as an anxiety reduction measure (cf. Farabee, Ramsey, & Cole, 1993; Stanley & Martin, 1968). Evidence from nonverbal communication research also suggests that anxious individuals fidget more than their nonanxious counterparts (Wenzel, Graff-Dolezal, Macho, & Brendle, 2005), justifying the expectation that anxiety-ridden liars would fidget more than truth-tellers.

Turning to the actual behaviors exhibited by liars, evidence from lie production again paints a conflicted picture. For example, Granhag and Strömwall (2002) report more self-manipulations, such as scratching or adjusting one's clothing, in liars, while Vrij and Winkel (1991) report fewer. In the case of gaze aversion, inconsistency is observed even within the same subject: In an analysis of the true and false statements produced by a convicted murderer during two separate police interviews, Vrij and Mann (2001) found that the subject showed more gaze aversion whilst lying than while truth-telling in one interview, but less in another. Vrij and Mann tentatively attribute this to the different styles of questioning employed by the officer in each interview, highlighting the potential mediating influence of interlocutor attitude on a liar's behavior (cf. Anolli & Ciceri, 1997).

Models of speaker deception

Within the framework of deception, two theoretical perspectives can each account for some of the divergent results in observed across studies. The first, the *cognitive hypothesis*, emphasizes the cognitive complexity associated with the act of lying. This hypothesis proposes that lying requires more mental effort, which in turn impacts a liar's behavior (Sporer & Schwandt, 2006; Vrij, 2000). The second, the *attempted control hypothesis*, focuses on the stereotypes of deceit and corresponding impression management measures employed by liars (Vrij, 1995). Under this hypothesis, speakers are aware that their behavior may reveal an intent to lie, and thus attempt to counteract potential exposure by controlling their speech and body language.

With regard to speech behavior, the cognitive hypothesis could explain a higher frequency of speech disturbances in liars, due to the increased mental load from having to construct a convincing lie. Vrij and Heaven (1999) systematically manipulated the complexity of the lie that speakers had to tell, and showed that the frequency of speakers' hesitations increased with lie complexity. This hypothesis receives further support from non-deception paradigms, which show that people engaged in cognitively complex tasks tend to speak slower and pause more (e.g., Goldman-Eisler, 1968; Kjellmer, 2003).

The attempted control approach, on the other hand, supports a decrease in liars' speech disturbances. For example, Villar, Arciuli, and Mallard (2012) observed that the speech of a convicted murderer contained fewer *ums* during false utterances, for statements produced in both private and public domains. This hypothesis is reinforced by evidence that speakers are able to regulate several aspects of their behavior when lying (DePaulo, Blank, Swaim, & Hairfield, 1992; R. Johnson, Henkell, Simon, & Zhu, 2008), and furthermore, that offering the right motivation (e.g., a monetary incentive) can reduce filled pause production to near zero levels (Boomer & Dittmann, 1964).[2]

The two hypotheses can each account for some of the divergent results in liars' nonverbal behavior. Speakers may experience difficulty formulating a lie spontaneously, which may manifest in the form of behaviors indicative of mental stress. Evidence from question-answer paradigms, for example, suggests that speakers avert their gaze due to the increased cognitive load associated with answering difficult questions (Doherty-Sneddon, Bruce, Bonner, Longbotham, & Doyle, 2002; Doherty-Sneddon & Phelps, 2005). Gesturing and hand movements have also been linked to cognitive load-reduction strategies that speakers employ as they

[2]Interestingly, research outside of deception suggests that speakers struggle to control some aspects of their speech, even when motivated to exercise control: Wardlow Lane and Liersch (2012) observed that speakers were actually more likely to disclose privileged information (in the form of a contrastive adjective) when given a monetary incentive to withhold this information. We thank an anonymous reviewer for pointing out this connection, which highlights that the motivation to deceive can, in different contexts, afford the speaker either increased control or alternatively irrepressible attention to the privileged information.

think about what to say (Goldin-Meadow et al., 2001). The cognitive hypothesis could thus explain why liars may avert their gaze or move their hands more as a by-product of the mental load associated with constructing a lie. On the other hand, liars, aware of the cue potential of their actions, may try to control these very behaviors to avoid being caught. The attempted control hypothesis may therefore explain why some researchers note that liars can come across as unusually rigid and inhibited as a result of overcontrolling their behavior (e.g., Vrij, 1995). This hypothesis is also often cited as a reason why studies consistently fail to identify reliable indicators of lying, as the discriminative potential of cues may diminish the more liars are able to effectively regulate their behavior (Granhag & Strömwall, 2002; cf. Buller, Comstock, Aune, & Strzyzewski, 1989).

Eliciting lies

There are many ways a speaker can deceive. Deception is broadly construed as an attempt to instill a false belief in someone (Hala et al., 1991)—this may range from a simple exaggeration of fact to blatant falsehood (see Knapp & Comaden, 1979 for a review of different forms of deception). The stereotypical form of deception is lying. This refers to the an act by which a speaker makes a false statement in order to instill a false belief in a listener (e.g., Talwar, Gordon, & Lee, 2007). Another way a speaker can deceive is by misleading—the act of leading someone to believe the opposite of what appears to be true on the surface. A misleading speaker can thus utter a simple falsehood (a lie that states the opposite of what is literally true), or else a truth whose intention is to lure a (distrustful) listener into thinking they are being lied to (double bluffing; cf. Happé, 1994).

In the current study, we employ an interactive game paradigm in which speakers are motivated to outwit a listener through misleading them. This serves as a useful way to elicit lies via 'free choice', as speakers are motivated to produce both true and false statements to advance in the game. Past researchers have criticized the tendency for studies to rely on cued lying paradigms, where speakers are directed to lie or tell the truth by means of a color or some other form of cue (e.g., Burgoon & Floyd, 2000). While this has the advantage of permitting a more balanced design, for example allowing for even numbers of true and false statements, such "instructed lies" may be problematic as they likely invoke different processes from those produced under the speaker's own volition. This issue was addressed in a recent neuroimaging study by Sip et al. (2010) which utilized a game paradigm where participants made truthful or false claims about a dice throw at will. Sip et al. observed that in contrast to previous cued-lying studies, false claims were not associated with activity in the dorsolateral prefrontal cortex (DLPFC). The authors attribute this to the fact that their task did not involve decision-making at the level of selecting appropriate responses in the current context, a process which typically invokes activity in the DLPFC (cf. Frith, 2000). This finding emphasizes the importance of allowing participants free choice on whether and when to lie in experiments.

Relying on instructed lies may also undermine a speaker's motivation to lie convincingly, raising doubts about the authenticity of the lie produced. The use of lies produced by unmotivated speakers in subsequent lie perception experiments further raises the issue of whether listeners are indeed discriminating between naturally produced truths and lies. DePaulo et al. (2003) compared studies in which speakers were offered inducements to succeed at lying to those which offered no special motivation, and found that cues were more pronounced in speakers that were motivated. In a similar vein, C. F. Bond and DePaulo (2006) observed across 20 studies that lies produced by motivated speakers were easier to classify than those produced by unmotivated speakers. These results highlight a role for motivation in the act of constructing a lie.

From a comprehension perspective, a related concern stems from a tendency for studies to elicit post-hoc judgments, where participants are asked to make a truth/lie discrimination judgment based on audio or video recordings of speakers (e.g., C. L. Hart et al., 2009). However, this approach removes listeners from the immediacy and interactivity that characterize a typical act of deception. Evidence from joint action research highlights that a listener's interpretation is closely linked to the act of interacting with another; for example, overhearers in the director-matcher task performed more poorly than matchers, who were actively engaged in conversation with the director (Schober & Clark, 1989). Listeners have also been found to follow instructions more accurately when produced in a dialogue rather than a monologue setting (Fox Tree, 1999). Hence, it seems plausible that listeners judging a recording of a speaker out of context would assess the speaker's veracity differently than when participating in the interaction.

We address these issues by employing an interactive game paradigm in which speakers lie at will, and listeners judge the veracity of each utterance in real-time.

Experiment

The current study is an exploratory investigation on speech and nonverbal cues to lying, with the aim of identifying which cues are produced during lies, and whether the same cues are used by listeners to infer when lying has occurred. The experiment was designed as a two-person competitive game, in which speakers produced a mix of true and false utterances about the location of a reward, while listeners made an implicit judgment of each utterance's veracity by attempting to discover the true location of the reward. As motivation to lie has been implicated as a moderator of liars' behavior (DePaulo et al., 2003), we also attempted to vary speakers' motivation by presenting two levels of reward—gold and silver coins, with different point values for each. We coded the speech and gestures produced by speakers for nineteen potential cues to lying. We employed a process of exploratory modeling based on the Akaike Information Criterion (AIC; Akaike, 1973) to determine which behavioral measures were reliable in predicting speakers' veracity, and listeners' judgments of veracity. By comparing the speech and nonverbal cues that speakers produced to those that influenced listeners' judgments, we made an explicit comparison of speakers' behavior and listeners' expectations surrounding the cues to lying.

6.2.3 Method

Participants

Twenty-four pairs of same-sex, native-British-English speakers took part in the study. Participants were all right-handed mouse users who reported no speech or hearing disorders. All provided informed consent in accordance with the University of Edinburgh Psychology Research Ethics Committee guidelines (reference no.: 214/1415-1).

Within each dyad, one member was assigned the role of Speaker (the liar) and the other the Guesser (the lie detector). All dyads were unacquainted prior to the study. Participants received £4 or course credits in exchange for participation. The winner of each dyad received an additional £1 cash reward.

Materials and design

The stimuli consisted of 96 black-and-white line drawings of objects, presented in fixed pairs across 48 trials. Forty-eight original images were drawn from Snodgrass and Vanderwart's (1980) data set. Slight modifications were then made to create a visually-related object-pair for each image (e.g., a camel with one/two humps). This was done with the aim of eliciting complex noun phrases from Speakers when naming an object in order to provide longer utterances for analysis. Each image pair thus consisted of two objects which were visually-related.

On each trial, one object within the pair was the target behind which the treasure was hidden, while the other served as a distractor. These were distinguished on the Speaker's display by a pile of coins or a pile of dirt behind each object respectively (see Fig 6.1). The coins were either gold (worth 20 points) or silver (worth 5 points). Eight lists were created, counterbalancing the role of each object within each pair (target or distractor), position of the target (left or right), and the type of treasure associated with the target (gold or silver) across all 96 objects. The order of presentation of image pairs was randomized across dyads.

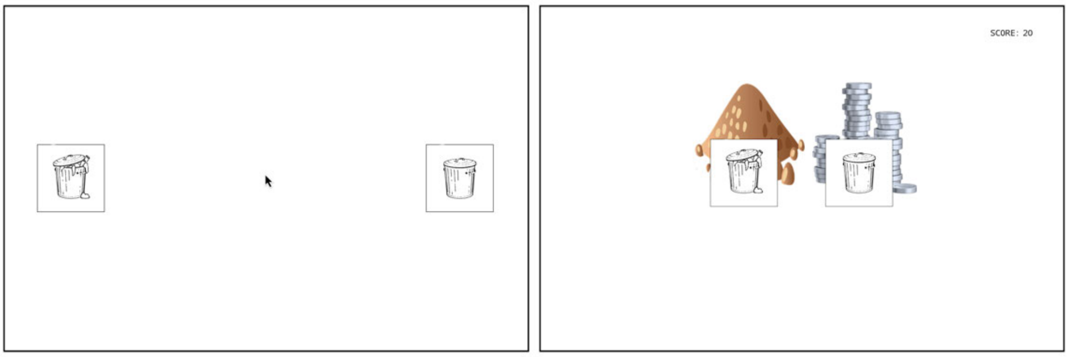


Figure 6.1: Example trial of the Guesser's display (left) and Speaker's display (right)

Procedure

The roles of Speaker and Guesser were assigned at the start by drawing lots. Speakers were instructed to describe the location of the treasure on each trial to their partner, with the aim of misleading them into looking for the treasure in the wrong location. They were given no additional guidance, other than that they were free to indicate the false object if they wished. Their ultimate goal was therefore to mislead the Guesser—they could do this either through lying (describing the false object with the expectation that the Guesser would believe them) or telling the truth (describing the correct object with the expectation that the Guesser interpret it as a false claim). Guessers were instructed to click on the object that they believed concealed the treasure, with the knowledge that their partner might be lying to them. Both players were present at the same time for the instructions, and were thus aware of the role and motivation of the other.

The experiment was presented using OpenSesame Version 2.9.5 (Mathôt et al., 2012) on 13 in. Apple Macintosh laptops. The Speaker and Guesser sat facing each other at diagonally opposite ends of a 24 x 36 in. table. This allowed positioning of video cameras at head level in front of each participant (see Fig 6.2).

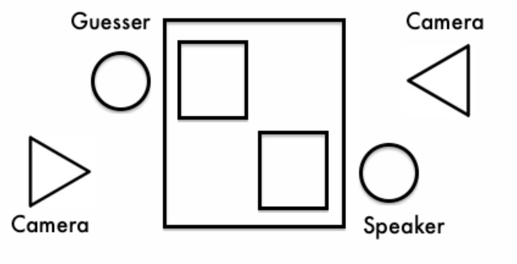


Figure 6.2: Diagrammatic setup of experiment.

Each trial began with a 1000 ms central fixation cross. This was then replaced by the image pair on each player's screen. On Guessers' displays, the objects were centered vertically and positioned horizontally left and right on the screen. The mouse cursor appeared at the center at the same time as the images. On Speakers' displays, the objects were set closer together to discourage Guessers from relying on Speakers' gaze locations (see Fig 6.1).

Once a click had been detected on an object, both players saw a message indicating whether the Guesser had found the treasure. Guessers received points upon successfully locating the treasure; Speakers received points when the Guesser chose the wrong object. Players' scores were cumulative over the course of the experiment. To incentivize dyads to compete against each other, a £1 cash reward was offered to the player with the higher score at the end.

A 3-trial practice session preceded the main experiment to familiarize dyads with the game, after which both players had the opportunity to ask questions before the game began. Trial progression was controlled from a separate computer by the experimenter, who initiated new trials when dyads were ready to continue after each one.

Participants' facial and upper body gestures were filmed during the experiment. Speech was recorded separately using a Zoom H4N digital recorder. This provided us with audio and video recordings of the dialogue for each Speaker and Guesser pair from the beginning of each trial up until the point where Guessers clicked on an object. Guessers' clicks on each trial (referent or distractor) were recorded, which provided an indirect indication of whether they judged the Speaker to be lying or telling the truth. Guessers' mouse coordinates were also sampled at 50Hz, with the aim of analyzing the trajectory of their decision-making in response to Speaker cues; however, due to the difficulty in mapping their mouse movements to Speakers' utterances, these were not analyzed further.

6.2.4 Results

Data

Speech from each trial was transcribed and annotated in Praat (Boersma & Weenink, 2013). Speakers' utterances on each trial were coded as truths (participant told the truth) or lies (participant lied about the treasure's location or the type of treasure—gold or silver coins). Guessers' responses were correspondingly coded as truth or lie judgments based on whether they clicked on the referent (object the speaker named as concealing the treasure) or the distractor.

Trials on which the Speaker was inconsistent in their commitment to the treasure's location were excluded from analyses (0.3% of all utterances) since it was impossible to determine on these trials whether the Speaker intended to lie or tell the truth from the outset. Additionally, as the experiment was interactive, Guessers did not always immediately click on an object following the Speaker's description. In cases of additional dialogue, such as when Guessers asked for clarification on an ambiguous description or attempted to catch Speakers out at lying (18.9% of trials), only the Speaker's initial utterance associated with the treasure's location was used for analysis.

The final dataset comprised 1,149 recorded utterances produced by 24 speakers. Video data from one participant was lost due to operator error, hence the final video dataset consisted of recordings of 1,101 trials from 23 speakers. Of the recorded utterances, Speakers were truthful 53.9% of the time (SE = 1.9) while Guessers judged 55.8% to be truthful (SE = 2.1). These figures are in line with the general trend in lie production and lie perception studies, which often note a global bias toward telling or expecting the truth (Vrij, 2000). The mean truth-lie discrimination accuracy for Guessers was 48.0% (SE = 1.4), with a 53.5% accuracy when Speakers were telling the truth and a 41.7% accuracy for Speakers' deceptive utterances. This difference is again unsurprising given the overall tendency for Guessers to perceive utterances as truthful (cf. C. F. Bond & DePaulo, 2006).

Annotation of utterances

The transcribed utterances were annotated for disfluencies by the first author and each disfluency was labeled for type. The following types of disfluency were identified: filled pauses, silent pauses, repetitions, restarts, substitutions, additions, and prolongations. To assess the reliability of the annotations, 20% of the speech data was randomly extracted and coded independently by a second coder. Both coders were blind to whether the Speaker referenced the treasure's true location in their utterance, and whether the Guesser selected the object described by the Speaker in their response. Appendix A provides the kappa statistics for interrater agreement between the two coders.

Table 6.1 presents the correlations between Speakers' truths, Guessers' perception of utterances as truths, and the seven disfluency types, along with three continuous speech measures (described below). Since some types of disfluency accounted for very few observations in the dataset (see Appendix A), the disfluencies were collapsed into four main categories. These were identified based on similar classification systems employed by existing studies (e.g., Hartsuiker & Notebaert, 2010; Merlo & Mansur, 2004; Shriberg, 1996). The categories include: 1) pauses, both filled (e.g., *uh*, *um* or *mm*) and silent, 2) repetitions, 3) repairs, where a verbalization was interrupted and restarted or corrected with a

Table 6.1: Correlations between Speakers' truths, Guessers' perception of utterances as truths, and individual speech variables.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Truths	1.00											
2. Perception of truths	0.08	1.00										
3. Filled pauses	0.12	-0.09	1.00									
4. Silent pauses	0.11	-0.17	-0.33	1.00								
5. Repetitions	0.01	0.01	-0.25	-0.42	1.00							
6. Restarts	0.19	-0.06	-0.30	-0.42	-0.20	1.00						
7. Substitutions	-0.07	-0.10	-0.07	-0.28	-0.21	0.03	1.00					
8. Additions	-0.12	0.11	-0.18	-0.24	0.03	-0.13	-0.19	1.00				
9. Prolongations	-0.01	-0.09	-0.17	-0.61	-0.19	0.00	0.18	-0.21	1.00			
10. Utterance duration	-0.05	0.07	0.31	0.56	0.18	0.28	0.07	0.08	0.47	1.00		
11. Silent pause duration	-0.03	0.09	0.17	0.59	0.23	0.17	0.05	0.07	0.35	-0.55	1.00	
12. Speech syllable rate	0.08	-0.06	-0.23	-0.68	-0.13	-0.13	-0.01	-0.03	-0.51	0.64	0.59	1.00

Note. Correlations were tetrachoric for associations between binomial variables (1–9); Pearson's for associations between continuous variables (10–12); and point-biserial for associations between binomial and continuous variables. All correlations were conducted at the observation level and do not take participant or item dependencies into account.

substitution or addition, and 4) prolongations. Table 6.2 provides examples of the disfluencies in each category.

Table 6.2: Disfluency categories and examples from data

Disfluency category	Example
Pause	behind **um** the banana that's not peeled
	behind the camel with **(0.32)** two humps
Repetition	behind the- **the** cut cake
Repair	**the money is th-** behind the one with the big tail fin
	behind the necklace which has beads coming- **falling** off it
	behind the open- **more** open book
Prolongation	behind **thee** leaf that looks like the ace on a pack of cards

The following continuous measures were also extracted from each utterance: duration of utterance; total silent pause duration within the utterance; and speech syllable rate (the number of perceptually salient syllables per second of speech). Speech onset latency was also measured but is not reported here as Speakers did not always begin with task-related speech (e.g., commenting on the stimuli or making other irrelevant observations). A general combined measure of speech rate was computed by extracting the first factor of a principal components analysis (PCA) on the three measures of utterance duration, silent pause duration, and speech syllable rate. This component had an eigenvalue of 2.19 and explained 73% of the variance. The PCA was conducted in R (version 3.2.4; R Core Team, 2017) using the FactoMineR library (Lê, Josse, & Husson, 2008).

Video recordings of Speakers on each trial were annotated in Elan (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) for their nonverbal behavior. The following gestures were identified: head movements, hand movements, body movements, shoulder movements, lip/mouth movements, eyebrow movements, smiling/laughter, and eye contact.[3] Videos for trials in which the Speaker was inconsistent regarding their commitment to the treasure's location were excluded. Only gestures produced during the duration of the Speaker's utterance were annotated. As with the speech data, 20% of the video data was extracted and coded independently by a second coder. Both coders were blind to the Speaker's veracity and the Guesser's judgment of veracity for each video clip. Appendix A provides the kappa statistics for interrater agreement between the two coders.

Table 6.3 presents the correlations between Speakers' truths, Guessers' perception of utterances as truths, and the eight gesture variables. The gestures were categorized as one of three main categories, identified based on Ekman and

[3]Gaze aversion is typically analyzed as a behavioral correlate of lying. However, due to the setup of our study which encouraged participants to look at their screen more than their interlocutor, gaze aversion was the norm, hence annotation took into account instances where eye contact was made.

Table 6.3: Correlations between Speakers' truths, Guessers' perception of utterances as truths, and individual gestures.

	1	2	3	4	5	6	7	8	9	10
1. Truths	1.00									
2. Perception of truths	0.08	1.00								
3. Head movements	-0.07	-0.01	1.00							
4. Hand movements	-0.12	0.08	-0.22	1.00						
5. Body movements	-0.10	-0.09	-0.11	-0.08	1.00					
6. Shoulder movements	0.12	0.11	-0.16	-0.07	0.00	1.00				
7. Lip/mouth movements	0.11	-0.10	-0.22	0.16	-0.10	-0.25	1.00			
8. Eyebrow movements	0.01	0.03	-0.41	0.12	-0.08	-0.12	-0.33	1.00		
9. Smiles/laughter	0.07	0.16	-0.27	-0.17	-0.30	-0.02	-0.15	-0.17	1.00	
10. Eye contact	-0.07	0.01	-0.37	-0.57	-0.15	0.10	0.19	-0.11	-0.30	1.00

Note. All correlations were tetrachoric. Correlations were conducted at the observation level and do not take participant or item dependencies into account.

Friesen's (1969) system of classifying nonverbal behavior. The categories are: 1) adaptors, encompassing self-oriented movements performed with little awareness and no message intent, 2) illustrators, defined as gestures designed to supplement or modify speech, and 3) affect displays, defined as gestures (primarily facial expressions) that function to convey specific emotions. A fourth category, eye contact, was used to represent instances when the Speaker made eye contact with the Guesser. Table 6.4 presents specific examples of gestures in each category.

Analysis

Statistical analyses were carried out in R using the lme4 package (Bates et al., 2014). We were interested in how well each of the speech and gesture categories predicted whether a) Speakers were telling the truth or lying, and b) Guessers perceived Speakers to be telling the truth or lying. Logistic mixed-effects regression models were used to model the outcome variables of Speakers' veracity (truth

Table 6.4: Gesture categories and examples from data

Gesture category	Example
Adaptor	Hand movements such as scratching one's head, adjusting one's clothing, clasping one's hands etc.
	Body movements such as rocking forwards, backwards or sideways
	Postural adjustments such as slumping or straightening one's back
Illustrator	Hand movements such as chopping motions to indicate a sliced carrot
	Head movements such as a head shake to indicate a tree with no fruit on it
Affect display	Eyebrow movements such as raised eyebrows to demonstrate surprise or furrowed brows to express concentration
	Mouth movements such as pursed lips to indicate thought
	Smiling or laughing during the utterance
Eye contact	Raising eyes from the screen to make eye contact with the Guesser

or lie) and Guessers' response (truth or lie, based on whether they clicked on the referent or distractor) on each utterance.

Due to the exploratory nature of the study and our goal of identifying variables of interest among a large set of predictors, we used the Akaike Information Criterion (AIC; Akaike, 1973) in a process of model evaluation to determine the best model given the data. AIC is a tool for comparing multiple models on a given outcome, allowing for comparison of non-nested models which cannot be handled by traditional hypothesis testing. AIC model ranking provides a useful method for evaluating the relative importance of multiple potential predictors (cf. T. W. Arnold, 2010), as is often the case in exploratory studies (e.g., Pagano &

Arnold, 2009; Snipes & Taylor, 2014).

AIC estimates information loss based on Kullback-Leibler (K-L; 1951) information quantity, which can be seen as the discrepancy between the model and full reality (Burnham, Anderson, & Huyvaert, 2011; Wagenmakers & Farrell, 2004). AIC model selection aims to minimize K-L information loss: Hence, smaller values, which denote less information loss, are preferred. Accordingly, the 'best' model among a set of competing models can be determined by selecting the model with the lowest AIC value (AIC$_{min}$).

The AIC value of a model is defined as $-2log(\mathcal{L}) + 2K$, where $log(\mathcal{L})$ is the maximized log-likelihood of the model and K is the number of estimable parameters (Burnham & Anderson, 2002). For small sample sizes, Hurvich and Tsai (1989) recommend a sample correction to the AIC:

$$\mathrm{AICc} = \mathrm{AIC} + \frac{(2K(K+1))}{(n-K-1)},$$

where n is the sample size and K and AIC are as defined above. It should be noted that raw AIC (and AICc) values on their own are meaningless; it is the difference in scores between models, the Δs, that provide a point for comparison. The difference in AIC for a given model i with respect to the best model is calculated by $\Delta\mathrm{AIC} = \mathrm{AIC}_i - \mathrm{AIC}_{min}$.

Our analyses utilized an AICc-based approach to select the best-supported model from a set of competing models designed to explain the outcomes of Speaker veracity and Guesser response. We conducted separate analyses to examine the effect of speech and gesture categories on each dependent variable. We first determined for each dependent variable a set of candidate models. This comprised all possible additive combinations of the set of predictors for the given outcome. For the speech models these included the four disfluency categories and speech rate, yielding a total of $2^5 = 32$ potential models, including a null (intercept-only) model. For the gesture models these included the three gesture categories and eye contact, yielding a total of $2^4 = 16$ potential models, again including the null. All models included random intercepts for participants and items (defined as the target image concealing the treasure on that trial).[4] All predictors other than speech rate were binary, and no distinction was made between one or more occurrences of the behavior during the utterance.

For each candidate model within a set, we calculated a) its AICc value; b) its AICc difference with respect to the best model (ΔAICc); c) its AICc weight (w_i), which provides a measure of the conditional probability of the model (Akaike, 1978; Wagenmakers & Farrell, 2004); and d) its evidence ratio (ER_i), which represents the strength of evidence of favoring the best model over that model (Wagenmakers & Farrell, 2004). We also computed e) the cumulative AICc weight ($\sum w_i$) for individual parameters by summing the AICc weights across all models including that variable (Tables 6.5 and 6.6). This provides a strength of evidence measure for each parameter (cf. T. W. Arnold, 2010), and is scaled between 0 (weakest) and 1 (strongest). Formulae used to derive measures c, d and e are given in Appendix B.

When evaluating which speech or gesture categories were reliable in predicting an outcome, we considered the model with the lowest AICc, as well as candidate models with a ΔAICc of less than 2 with respect to that model (cf. Burnham & Anderson, 2002). Appendix C provides a complete list of all the models in each candidate set along with their associated AICc ranking, AICc weight and evidence ratio, as well as the final model output for the best-supported model (as determined by AICc_{min}) from each candidate set.

To explore the effect of motivation to lie on Speakers' behavior, secondary analyses were conducted to model motivation as a predictor for each behavioral measure. As we were interested in the (potential) moderating effect of motivation, this analysis was limited to behavioral measures which emerged as significant in

[4]The inclusion of random slopes led to non-convergence for many of the models with 3 or more predictors, hence slopes were omitted from all models to maintain a consistent random effects structure.

predicting Speakers' veracity. The dependent variables were the behaviors of interest, with motivation as a binary (high/low) fixed effect. All models included random intercepts and random slopes for participants and random intercepts for items.

Speech cues

Table 6.5 presents the cumulative AICc weights of the speech parameters used to model Speaker veracity and Guesser response.

Table 6.5: Cumulative AICc weights ($0 \leqslant \sum w_i \leqslant 1$) of speech model parameters for Speaker veracity and Guesser response

	$\sum w_i$	
Model parameter	Speaker veracity	Guesser response
pauses	0.61	0.79
repetitions	0.27	0.33
repairs	0.56	0.64
prolongations	0.43	0.36
speech rate	0.43	0.33

The best-supported model in estimating the effect of the speech variables on predicting Speakers' veracity was the model containing only pauses. This model had an AICc weight of 0.108, indicating that it accounted for 10.8% of the total weight of all models in the candidate set, and ranked 0.70 AICc units above the second best-supported model. The model was 1.42 times more likely than the next best model to be the most parsimonious for the data, as indicated by the evidence ratio of the latter.

Model coefficients showed a positive relationship between pauses and veracity: Speakers were 1.3 times more likely to be telling the truth than lying when their utterance contained a pause, $\beta = 0.26$, SE = 0.13, $p = .04$ ($e^{0.26} = 1.30$). Correlations between Speakers' truths and individual speech cues show that both filled and silent pauses correlated positively with truth-telling (Table 6.1), suggesting that the relationship between pauses and veracity was driven by both forms of pauses.

Of the 32 candidate models, 8 were within 2ΔAICc of the best model and hence can be interpreted as competitive in predicting the given outcome. Of these 8 models, 6 incorporated pauses. Pauses also had the highest cumulative AICc weight of the 5 variables (Table 6.5), lending support to the influence this variable had in predicting outcome. The remaining 2 supported models incorporated only speech rate (ΔAICc = 1.36) and prolongations and speech rate (ΔAICc = 1.34) respectively.

For Guessers, the best-supported model in estimating the effect of the speech variables on response judgment was the model containing only pauses. This model accounted for 15.1% of the total weight of all models in the set, as indicated by its AICc weight, and ranked 0.26 AICc units above the second best-supported model. The model was 1.14 times more likely to be considered the best model for the given data than the second best model, based on the evidence ratio of the latter.

In contrast to Speakers, the model coefficients for Guessers showed a negative relationship between pauses and truth perception: Guessers were 0.67 times as likely to click on the referent (and therefore more likely to click on the distractor) when the utterance contained a pause, $\beta = -0.39$, SE = 0.13, $p < .01$. In other words, pauses were more likely to be associated with lying by Guessers. Correlations between Guessers' truth perception and individual speech cues show that both filled and silent pauses correlate negatively with truth perception (Table 6.1), suggesting that the relationship between pauses and Guesser response was driven by both types of pauses.

Of the candidate models, 6 were within 2ΔAICc of the best model. All 6 competitive models incorporated pauses. The influence of pauses on Guessers' responses is also supported by the cumulative AICc weights, which show that pauses had the highest weight of all 5 speech variables (Table 6.5).

Together, these results highlight the role of pauses in predicting Speaker veracity and Guesser response. Analysis of the best-supported model for each candidate set, the subset of models deemed competitive, and the cumulative AICc weights of the individual model parameters provide unified evidence in support of this variable. Model coefficients also indicate a difference in the direction of effect on each outcome variable: Although pauses were an index of truth-telling in Speakers, they were associated with lie judgments in Guessers.

Gesture cues

Table 6.6 presents the cumulative AICc weights of the gesture parameters used to model Speaker veracity and Guesser response.

Table 6.6: Cumulative AICc weights ($0 \leqslant \sum w_i \leqslant 1$) of gesture model parameters for Speaker veracity and Guesser response

	$\sum w_i$	
Model parameter	Speaker veracity	Guesser response
adaptors	0.81	0.27
affect displays	0.28	0.76
illustrators	0.36	0.30
gaze behavior	0.32	0.27

The best-supported model in estimating the effect of gestures on predicting Speakers' veracity was the model containing only adaptors. This model accounted for 26.3% of the total weight of all models in the set, as indicated by its AICc weight, and ranked 1.27 AICc units above the second best-supported model. The model was 1.88 times more likely to be the best model for the given data than the next best model, as indicated by the evidence ratio of the latter.

Model coefficients showed that Speakers' veracity varied with their production of adaptors: Speakers were 0.75 times as likely to be telling the truth (and therefore more likely to be lying) when their utterance was accompanied by an adaptor, $\beta = -0.29$, SE = 0.13, $p = .02$. Of the three gesture cues that constituted adaptors, hand and body movements correlated negatively with Speakers' truths while shoulder movements correlated positively (Table 6.3), suggesting that the relationship between adaptors and veracity was driven primarily by Speakers' hand and body movements.

Three of the 16 candidate models were within 2ΔAICc of the best model, all of which incorporated adaptors. Adaptors also had the highest cumulative AICc weight of all 4 gesture variables on Speaker veracity (Table 6.6).

For Guessers, the best-supported model for the effect of gestures on response was the model containing only affect displays. This model accounted for 28.1% of the total weight of all models in the set, as indicated by its AICc weight, and ranked 1.64 AICc units above the second best-supported model. The model was 2.27 times more likely to be the best model for the data than the next best model, as indicated by the evidence ratio of the latter.

Model coefficients showed a positive relationship between affect displays and truth perception: Guessers were 1.34 times more likely to click on the referent when the utterance was accompanied by an affect display, $\beta = 0.29$, SE = 0.14, $p = .04$. In other words, Speakers' affect displays were more likely to be associated with truth-telling by Guessers. Of the gestures that comprised affect displays, lip movements correlated negatively with truth perception while eyebrow movements and smiles/laughter correlated positively, suggesting that the relationship between affect displays and Guesser response was driven primarily by the latter two behaviors.

Of the 16 candidate models, 3 were within 2ΔAICc of the top model, all of which incorporated affect displays. Affect displays also had the highest cumulative AICc weight of all 4 gesture variables on Guessers' responses (Table 6.6).

Results from the gesture analyses thus highlight different variables in predicting Speaker veracity and Guesser response. Evidence from the best-supported model, the subset of competitive models, and the cumulative AICc weights from each analysis show that adaptors had the greatest influence in explaining veracity, while affect displays had the greatest influence in explaining response. Model coefficients reveal a difference in the direction of effect of each variable on its outcome: Adaptors were an index of falsehood in Speakers, while affect displays were interpreted as a sign of truth-telling by Guessers.

Effect of motivation on cue production

The analysis on the effect of motivation on Speaker's cue productions focused on behaviors that were found to significantly predict whether they were lying or telling the truth—pauses and adaptors. This analysis showed that neither behavior was found to vary with motivation (all ps $> .7$). Thus, Speakers' motivation to lie did not appear to affect their production of any cue behaviors which varied with the veracity of their utterances.

6.2.5 Discussion

In the present study, we investigated the production and perception of speech and nonverbal cues to lying. We were interested in which cues could be used to predict a speaker's veracity, and whether those same cues are used by listeners to judge a speaker's veracity. To that aim, we designed a task that elicited both true and false utterances from speakers, and a corresponding truth or lie judgment by a listener on each utterance. This allowed us to obtain judgments on utterances as participants interacted in real time, unlike previous studies which have employed post-hoc methods of judgment (e.g., C. L. Hart et al., 2009).

We observed an overall bias towards truthfulness in both Speakers and Guessers. This distribution aligns with the existing literature on deception, which highlights a tendency for speakers to tell the truth (Beňuš, Enos, Hirschberg, &

Shriberg, 2006), and a tendency for listeners to interpret utterances as truthful (Levine et al., 1999). Our analyses employed an information-theoretic approach to evaluate the reliability of multiple cue behaviors in explaining Speaker veracity and Guesser response. With respect to the cues that emerged as significant, we found a surprising mismatch between the cues that Speakers produced when lying, and those that Guessers attended to when judging veracity.

From a methodological perspective, our results show that AICc-based model selection can be a useful approach to exploratory psycholinguistic studies investigating a large number of potential predictors. The large number of predictors is often the case with research on disfluencies (e.g., Merlo & Mansur, 2004; Shriberg, 1996) or deception (e.g., Granhag & Strömwall, 2002; Vrij, Akehurst, & Knight, 2006), but can also extend more widely to areas such as discourse processing (e.g., Quek et al., 2000; White, 1989). For such studies, AICc modeling presents an alternative to traditional methods of analysis, which have often involved independent testing of individual predictors, leading to problems associated with multiple inference (cf. Curran-Everett, 2000).

As with any methodology, however, AICc has its drawbacks. In particular, the 2ΔAICc rule used to establish a subset of competitive models can be criticized as being an arbitrary cutoff, akin to the $p < .05$ significance rule in traditional null hypothesis testing; and indeed, various interpretations of ΔAIC (or ΔAICc) have been adopted by different researchers as evidence for model support (see Murtaugh, 2014, p.651, for a summary).

Where AICc improves upon other methodologies is the quality of the final model used to explain the data. Recall that the aim of AICc-based model selection is to find the best model that balances descriptive accuracy by adding more parameters with loss of generalizability from over-fitting. As such, the best-supported model can be seen as evidence of which variables among a set are informative in explaining an outcome and which are not (cf. Snipes & Taylor, 2014).

It is important to note that the results do not provide definitive evidence for one model over the others; rather, the AICc trends in each analysis should be taken as combined evidence to support conclusions drawn from the data. In using an AICc-based approach, Burnham et al. (2011) recommend taking into account model likelihoods, evidence ratios, and consideration of all candidate models in the set to form a more-informed overall inference. With that in mind, we computed AICc weights and evidence ratios for all models as well as cumulative AICc weights for individual model parameters. From our results, it is clear that these quantitative measures provide corroborative evidence in support of the variables that emerged as significant in each of the best-supported models.

Our exploratory modeling revealed a mismatch between the cues that reliably predicted Speakers' veracity and those that Guessers attended to in judging veracity. The best model for Speakers revealed pauses to be a significant predictor of veracity: Speakers were more likely to be telling the truth when their utterance contained a pause, either filled or silent. In contrast, although pauses also emerged as a significant predictor in the best Guesser model, this relationship was in the opposite direction to Speakers' behavior: Guessers interpreted Speaker's utterances containing pauses to be lies. This pattern in Guessers corroborates findings from existing lie perception research (Akehurst et al., 1996; Zuckerman, Koestner, & Driver, 1981), and paints a portrait of a liar as one embroiled in a state of difficulty, either due to the cognitive burden of having to formulate a lie (Vrij et al., 2000), or that of experiencing various negative emotions associated with the act of lying (Ekman, 2001).

The Speaker-Guesser disconnect parallels the pattern observed in our exploratory gesture modeling. In this case, the best Speaker and Guesser models highlight distinct cues respectively as predictors of veracity and judgment of veracity. While affect displays were associated with lying by Guessers, Speakers' veracity did not vary with their production of this cue; rather, it was their production of adaptors that betrayed an intent to lie. Guessers' interpretations of

Speakers' lack of affect displays as a sign of falsehood also aligns with existing research, which highlights facial cues such as decreased smiling or an unfriendly facial expression as indicators of lying (Vrij et al., 2006; Zuckerman, DePaulo, & Rosenthal, 1981), and further supports cognitive or emotional-based explanations in lie perception.

Although Guessers' beliefs about cues to lying appear driven by the cognitive hypothesis, Speakers' behavior may be better explained by the attempted control approach to deception. This is evident through the significantly fewer pauses they produced when lying—the very cue that Guessers attended to when inferring that lying had occurred. This may suggest that Speakers were sensitive to Guessers' expectations regarding disfluency and lying, and consequently attempted to suppress cue production to present a credible impression. However, this account appears counter to Speakers' nonverbal behavior, which demonstrates an increase in adaptors during lie production. The discrepancy between Speakers' verbal and nonverbal behavior is, however, in keeping with Ekman and Friesen's (1969) *leakage hierarchy hypothesis*, which posits that some channels of communication are harder for speakers to control than others. These are the so-called 'leaky channels' which in turn expose deceit. According to Ekman and Friesen, people should be more successful at monitoring their facial behavior (except for micro-expressions; cf. Ekman, 2001) when lying than monitoring their hands, feet or bodies, due to the social salience of facial expressions in communication (cf. Vrij et al., 2001).

Similarly, speech cues (with the exception of voice pitch) are held to be more controllable than many aspects of nonverbal behavior (Ekman, O'Sullivan, Friesen, & Scherer, 1991; Sporer & Schwandt, 2006). Beyond deception research, evidence further shows that disfluency production is under a speaker's conscious control, and under the right circumstances can be reduced or eliminated (Boomer & Dittmann, 1964; Broen & Siegel, 1972). Given these findings, it seems unsurprising that Speakers in our experiment showed a decrease in pause behavior but an increase in adaptors when lying—liars' inability to control all aspects of their behavior equally could have resulted in an overcontrol of their speech cues but a leakage in their body language (cf. DePaulo et al., 2003).

There is one other potential explanation for why disfluency indexes truth-telling in this task, which is that in some cases, such as in the present paradigm, truth-telling may be more difficult for speakers than lying. A prevailing assumption within the literature is that deception simply involves a straightforward lie where the speaker asserts a false claim (e.g., C. L. Hart et al., 2009; Vrij et al., 2001). However, the present paradigm involved deception on two levels—simple falsehood as well as misleading-via-truth-telling. The latter form is arguably more subtle and sophisticated, as it involves a speaker reasoning that the listener may disbelieve their (truthful) message (cf. Sutter, 2009). It is, therefore, possible that truth-telling with an intent to deceive actually places a greater mental load on speakers than straightforward lying. Under a cognitive hypothesis, this may lead to an increase in certain cue behaviors such as speech disturbances and adaptors. While a classification of what deception should encompass lies beyond the focus of the present study, we venture that current models should broaden their definition to consider different forms of deception and the mechanisms underlying each form.

Finally, neither behavioral measure which emerged as significant in predicting Speakers' veracity was found to vary with their motivation to lie. This result is difficult to reconcile with existing studies examining the effect of motivation on lie behavior, many of which report an increase in cue production as a function of liars' motivation (DePaulo & Kirkendol, 1989; DePaulo et al., 2003). However, this discrepancy may be due to several key differences between previous studies and ours.

Previous research on the modulating effect of motivation on a liar's behavior has typically characterized this variable as a speaker's motivation to produce a successful lie. As such, effects were measured based on listeners' success at detecting those lies, and consequently, cue leakage is assumed based on listeners' accurate classification of lies. This stands in contrast to our study, in which motivation was investigated more directly: We compared speakers' cue behaviors when they were more or less motivated, independent of the listeners' ultimate detection rate. Moreover, motivation is frequently implemented as a between-subjects variable, where participants are randomly assigned to a high or low motivation condition (e.g., DePaulo, Kirkendol, Tang, & O'Brien, 1988; DePaulo, Lanier, & Davis, 1983; Hancock, Woodworth, & Goorha, 2010). Meta-analytic studies investigating the link between lying and motivation similarly treat motivation as an experimental-level factor, comparing studies which offered speakers inducement for lie success to those which did not (C. F. Bond & DePaulo, 2006; DePaulo et al., 2003). This operationalization of motivation differs from our design, where motivation was varied within-subjects by using two levels of reward within each game. As real-life deception would typically see a liar's motivation fluctuate from context-to-context, closer exploration of how its effect varies within speakers would be a useful avenue for future research. Finally, superficial differences in tasks may have produced qualitatively different behavior in our speakers and those in earlier studies. In particular, the game setup of our experiment meant that Speakers' baseline level of motivation was likely high to begin with. This is confirmed by our conversations at debrief, which suggested that participants felt incentivized to succeed at the game and beat their opponent. As such, it is possible that any further impact from varying motivation within the game may have been too subtle to result in demonstrable differences in behavior.

Although the game paradigm we used may have obscured any effects of a motivation manipulation, we maintain that such a design has its advantages in deception research. Extending previous studies which have examined cues to lying from a unidirectional perspective, we were able to investigate the simultaneous production and perception of cues using an interactive framework. This enabled us to simulate authentic, real-world deception in a controlled, yet relatively natural context. By allowing Speakers to choose when to lie and when to tell the truth, we also avoided problems associated with cued-lying paradigms, where the directed nature of the task may undermine the authenticity of the lies (cf. Spence, Kaylor-Hughes, Farrow, & Wilkinson, 2008). This provides ecological validity to Speakers' utterances, their behavioral cues accompanying those utterances, and Guessers' responses to those cues. The results we observed suggest that a liar's behavior is influenced not only by the act of conceiving a lie, but by the expectations that listeners may have regarding the speaker's speech and gestures. Thus, the study highlights the importance of considering the interactive dimension in lie production and lie perception paradigms in order to contribute to a more complete understanding of the psychological dynamics that shape an act of deception.

6.2.6 Appendix A

Table 6.7: Descriptive statistics and Cohen's Kappa (κ) between the two coders for the individual speech and gesture variables.

	raw count	mean (SD)	κ
Speech variables ($n = 1,149$)			
Filled pauses	288	-	.95
Silent pauses	588	-	.97
Repetitions	55	-	.87
Restarts	109	-	.95
Substitutions	36	-	.95
Additions	12	-	1.0
Prolongations	334	-	.82
Utterance duration	-	3008.92 (1329.35)	-
Silent pause duration	-	651.65 (1080.5)	-
Speech syllable rate	-	3.82 (1.42)	-
Gestures ($n = 1,101$)		-	
Head movements	651	-	.76
Hand movements	280	-	.92
Body movements	377	-	.87
Shoulder movements	26	-	.85
Lip/mouth movements	85	-	.50
Eyebrow movements	242	-	.83
Smiles/laughter	156	-	.81
Gaze	130	-	.95

6.2.7 Appendix B

Formulae used to derive AICc weights and evidence ratios of models, and cumulative AICc weights of model parameters:

AICc weight: $w_i = \dfrac{exp(-\frac{1}{2}\Delta\text{AICc}_i)}{\sum\limits_{j=1}^{J} exp(-\frac{1}{2}\Delta\text{AICc}_j)}$

Evidence ratio: $ER_i = \dfrac{w_{best}}{w_i}$

Cumulative AICc weight of parameter p: $\sum w_i = \sum\limits_{i=1}^{P} w_i$,

where i denotes the current model, J is number of models in the candidate set, w_{best} is the AICc weight of the best model, and P is the number of candidate models in the subset of models containing parameter p.

6.2.8 Appendix C

Table 6.8: Summary of AICc results for models exploring the effect of speech variables on predicting Speaker veracity. Models are ranked according to differences in Akaike's Information Criterion corrected for small sample sizes (AICc).

Model[a]	AICc	ΔAICc	w_i[b]	ER_i[c]
y ∼ ps + ε	1583.04	0.00	0.108	1.00
y ∼ ps + pl + ε	1583.73	0.70	0.076	1.42
y ∼ ps + rp + ε	1583.91	0.88	0.070	1.55
y ∼ pl + sr + ε	1584.38	1.34	0.055	1.95
y ∼ sr + ε	1584.39	1.36	0.055	1.97
y ∼ ps + rp + pl + ε	1584.74	1.71	0.046	2.35
y ∼ ps + pl + sr + ε	1584.75	1.71	0.046	2.35
y ∼ ps + sr + ε	1584.82	1.78	0.044	2.44
y ∼ ps + rr + ε	1585.02	1.99	0.040	2.70
y ∼ 1 + ε	1585.04	2.00	0.040	2.72
y ∼ rp + ε	1585.05	2.01	0.040	2.73
y ∼ rp + sr + ε	1585.30	2.26	0.035	3.09
y ∼ rp + pl + sr + ε	1585.64	2.60	0.030	3.69
y ∼ ps + rr + pl + ε	1585.73	2.70	0.028	3.85
y ∼ ps + rp + sr + ε	1585.84	2.80	0.027	4.06
y ∼ ps + rr + rp + ε	1585.87	2.83	0.026	4.12
y ∼ ps + rp + pl + sr + ε	1586.12	3.07	0.023	4.64
y ∼ rr + pl + sr + ε	1586.32	3.28	0.021	5.16
y ∼ rr + sr + ε	1586.36	3.32	0.021	5.26
y ∼ ps + rr + pl + sr + ε	1586.68	3.64	0.018	6.18
y ∼ ps + rr + rp + pl + ε	1586.72	3.68	0.017	6.31
y ∼ ps + rr + sr + ε	1586.78	3.74	0.017	6.49
y ∼ pl + ε	1586.82	3.79	0.016	6.64
y ∼ rp + pl + ε	1586.83	3.80	0.016	6.67
y ∼ rr + ε	1587.04	4.01	0.015	7.41
y ∼ rr + rp + ε	1587.06	4.03	0.014	7.49
y ∼ rr + rp + sr + ε	1587.23	4.20	0.013	8.15
y ∼ rr + rp + pl + sr + ε	1587.55	4.52	0.011	9.57
y ∼ ps + rr + rp + sr + ε	1587.77	4.74	0.010	10.67
y ∼ ps + rr + rp + pl + sr + ε	1588.02	4.98	0.009	12.08
y ∼ rr + pl + ε	1588.82	5.79	0.006	18.06
y ∼ rr + rp + pl + ε	1588.85	5.82	0.006	18.32

[a] y = Speaker veracity, ps = pauses, rr = repetitions, rp = repairs, pl = prolongations, sr = speech rate, ε = error term (random intercepts for subjects and items)

[b] AICc weight (relative likelihood of model based on AICc value)

[c] Evidence ratio (likelihood of favoring the best model over the current model i)

Table 6.9: Model output for best-supported model for the effect of speech variables on Speaker veracity.

Model of pauses on Speaker veracity				
Fixed effects				
	Estimate	SE	z value	p value
(Intercept)	0.16	0.08	1.96	.05
pauses	0.26	0.13	2.01	.04
Random effects				
		Variance	SD	
participants	(Intercept)	0.05	0.23	
items	(Intercept)	0.10	0.31	
No. of obs: 1149, groups: participant, 24; item, 96				

Table 6.10: Summary of AICc results for models exploring the effect of speech variables on predicting Guesser response. Models are ranked according to differences in Akaike's Information Criterion corrected for small sample sizes (AICc).

Model[a]	AICc	ΔAICc	w_i[b]	ER_i[c]
y ∼ ps + ε	1569.21	0.00	0.151	1.00
y ∼ ps + rr + ε	1569.47	0.26	0.132	1.14
y ∼ ps + sr + ε	1570.80	1.60	0.068	2.22
y ∼ ps + pl + ε	1570.88	1.67	0.065	2.31
y ∼ ps + rr + ε	1570.92	1.71	0.064	2.35
y ∼ ps + rp + pl + ε	1571.07	1.86	0.060	2.54
y ∼ ps + rr + rp + ε	1571.07	1.86	0.059	2.54
y ∼ rr + rp + pl + ε	1571.33	2.12	0.052	2.89
y ∼ ps + rr + sr + ε	1572.38	3.17	0.031	4.88
y ∼ sr + ε	1572.44	3.24	0.030	5.05
y ∼ ps + rr + pl + ε	1572.57	3.37	0.028	5.38
y ∼ ps + rr + pl + sr + ε	1572.64	3.43	0.027	5.57
y ∼ ps + pl + sr + ε	1572.69	3.48	0.026	5.71
y ∼ rp + sr + ε	1572.72	3.51	0.026	5.79
y ∼ ps + rp + pl + sr + ε	1572.84	3.63	0.025	6.15
y ∼ ps + rp + sr + ε	1573.07	3.86	0.022	6.90
y ∼ rr + sr + ε	1574.04	4.83	0.013	11.20
y ∼ rp + ε	1574.07	4.86	0.013	11.36
y ∼ rp + pl + ε	1574.18	4.97	0.013	12.03
y ∼ rp + pl + sr + ε	1574.24	5.03	0.012	12.39
y ∼ ps + rr + rp + sr + ε	1574.28	5.08	0.012	12.67
y ∼ pl + sr + ε	1574.30	5.10	0.012	12.80
y ∼ ps + rr + rp + pl + ε	1574.41	5.20	0.011	13.47
y ∼ ps + rr + rp + pl + sr + ε	1574.60	5.40	0.011	14.85
y ∼ pl + ε	1575.72	6.51	0.006	25.94
y ∼ 1 + ε	1575.73	6.53	0.006	26.17
y ∼ rr + rp + sr + ε	1575.92	6.71	0.005	28.68
y ∼ rr + rp + pl + sr + ε	1575.95	6.75	0.005	29.19
y ∼ rr + rp + ε	1575.97	6.76	0.005	29.38
y ∼ rr + pl + sr + ε	1576.01	6.81	0.005	30.04
y ∼ rr + pl + ε	1577.67	8.46	0.002	68.82
y ∼ rr + ε	1577.72	8.52	0.002	70.71

[a] y = Guesser response, ps = pauses, rr = repetitions, rp = repairs, pl = prolongations, sr = speech rate, ε = error term (random intercepts for subjects and items)

[b] AICc weight (relative likelihood of model based on AICc value)

[c] Evidence ratio (likelihood of favoring the best model over the current model i)

Table 6.11: Model output for best-supported model for the effect of speech variables on Guesser response.

Model of pauses on Guesser response				
Fixed effects				
	Estimate	SE	z value	p value
(Intercept)	0.24	0.78	3.06	$< .01$
pauses	-0.39	0.13	-2.95	$< .01$
Random effects				
		Variance	SD	
participants	(Intercept)	0.06	0.24	
items	(Intercept)	0.00	0.00	
No. of obs: 1147, groups: participant, 24; item, 96				

Table 6.12: Summary of AICc results for models exploring the effect of gesture variables on predicting Speaker veracity. Models are ranked according to differences in Akaike's Information Criterion corrected for small sample sizes (AICc).

Model[a]	AICc	ΔAICc	w_i[b]	ER_i[c]
$y \sim ad + \varepsilon$	1516.03	0.00	0.264	1.00
$y \sim ad + il + \varepsilon$	1517.30	1.27	0.140	1.88
$y \sim ad + ey + \varepsilon$	1517.69	1.66	0.115	2.29
$y \sim ad + af + \varepsilon$	1517.94	1.91	0.102	2.60
$y \sim ad + il + ey + \varepsilon$	1518.68	2.64	0.070	3.75
$y \sim 1 + \varepsilon$	1519.14	3.11	0.056	4.73
$y \sim ad + af + il + \varepsilon$	1519.28	3.25	0.052	5.08
$y \sim ad + af + ey + \varepsilon$	1519.53	3.50	0.046	5.75
$y \sim il + \varepsilon$	1520.30	4.27	0.031	8.47
$y \sim ey + \varepsilon$	1520.45	4.42	0.029	9.13
$y \sim ad + af + il + ey + \varepsilon$	1520.61	4.58	0.027	9.87
$y \sim af + \varepsilon$	1521.15	5.12	0.020	12.91
$y \sim il + ey + \varepsilon$	1521.23	5.20	0.020	13.49
$y \sim af + il + \varepsilon$	1522.32	6.29	0.011	23.19
$y \sim af + ey + \varepsilon$	1522.42	6.39	0.011	24.44
$y \sim af + il + ey + \varepsilon$	1523.25	7.22	0.007	36.93

[a] y = Speaker veracity, ad = adaptors, af = affect displays, il = illustrators, ey = eye contact, ε = error term (random intercepts for subjects and items)

[b] AICc weight (relative likelihood of model based on AICc value)

[c] Evidence ratio (likelihood of favoring the best model over the current model i)

Table 6.13: Model output for best-supported model for the effect of gesture variables on Speaker veracity.

Model of adaptors on Speaker veracity					
Fixed effects					
		Estimate	SE	z value	p value
(Intercept)		0.18	0.08	2.11	.04
adaptors		−0.29	0.13	−2.26	.02
Random effects					
			Variance	SD	
participants	(Intercept)		0.08	0.28	
items	(Intercept)		0.06	0.23	
No. of obs: 1101, groups: participant, 23; item, 96					

Table 6.14: Summary of AICc results for models exploring the effect of gesture variables on predicting Guesser response. Models are ranked according to differences in Akaike's Information Criterion corrected for small sample sizes (AICc).

Model[a]	AICc	ΔAICc	w_i[b]	ER_i[c]
y ∼ af + ε	1508.12	0.281	1.00	
y ∼ af + il + ε	1509.76	0.124	2.27	
y ∼ ad + af + ε	1510.08	0.105	2.67	
y ∼ af + ey + ε	1510.08	0.105	2.67	
y ∼ 1 + ε	1510.36	0.092	3.06	
y ∼ ad + af + il + ε	1511.72	0.046	6.06	
y ∼ af + il + ey + ε	1511.76	0.045	6.20	
y ∼ ad + af + ey + ε	1512.06	0.039	7.19	
y ∼ il + ε	1512.24	0.036	7.87	
y ∼ ey + ε	1512.37	0.033	8.39	
y ∼ ad + ε	1512.37	0.033	8.39	
y ∼ ad + af + il + ey + ε	1513.73	0.017	16.60	
y ∼ il + ey + ε	1514.26	0.013	21.55	
y ∼ ad + il + ε	1514.26	0.013	21.57	
y ∼ ad + ey + ε	1514.39	0.012	23.01	
y ∼ ad + il + ey + ε	1516.28	0.005	59.21	

[a] y = Guesser response, ad = adaptors, af = affect displays, il = illustrators, ey = eye contact, ε = error term (random intercepts for subjects and items)

[b] AICc weight (relative likelihood of model based on AICc value)

[c] Evidence ratio (likelihood of favoring the best model over the current model i)

Table 6.15: Model output for best-supported model for the effect of gesture variables on Guesser response.

Model of affect displays on Guesser response				
Fixed effects				
	Estimate	SE	z value	p value
(Intercept)	0.23	0.90	2.57	.01
affect displays	0.29	0.14	2.05	.04
Random effects				
		Variance	SD	
participants	(Intercept)	< 0.01	< 0.01	
items	(Intercept)	< 0.01	< 0.01	
No. of obs: 1101, groups: participant, 23; item, 96				

6.2.9 Analysis of mouse movements

Guessers' mouse coordinates were tracked during the experiment with the aim of analysing the time course of their veracity judgements in response to Speakers' utterances. These were not reported in the manuscript of the main study for reasons of space as well as technical difficulties encountered in interpreting the data accurately. However, to provide readers with a full picture, in this section we present the analyses and results based on the mouse movement data.

We first established the point of disambiguation for each transcribed utterance in the dataset. This was defined as the earliest moment at which the Guesser could reasonably identify which of the two objects the Speaker was referencing based on their unfolding description. For the images in Fig. 6.1, for example, the point of disambiguation was determined as follows (indicated in boldface) in each of the utterances below:

(10) a. behind the **open** bin
b. behind the bin **full** of rubbish
c. behind the bin with**out** the rubbish
d. behind the bin that is not **overflowing** with dirt

Guessers' mouse movements were time-locked to the onset of this point for analysis. All annotations were conducted by the first author of the main study, who was blind to whether the Speaker produced a truth or a lie and whether the Guesser believed the utterance to be a true or false description of the treasure's location. Trials on which the point of disambiguation was indeterminable were not annotated and were excluded from analysis. These included cases of additional dialogue (e.g., when Guessers asked for clarification), such as in (11), or when Guessers attempted to catch Speakers out at lying (e.g., by asking leading questions), such as in (12):

(11) S: the treasure is behind the candle that isn't f- very melted
G: isn't very melted
S: yeah the like fresh candle

(12) S: ok the treasure is behind the spotty tie
G: are you telling the truth
S: you're just going to have to work that out aren't you

Post exclusion, the final dataset contained 970 utterances. To analyse Guessers' mouse movements on each trial, we followed the same processing procedure as in our earlier experiments to calculate proportion-of-movement measures toward each object (referent or distractor) for each 20 ms time bin (see section 4.2.5). For each bin, we first computed the horizontal distance travelled by the mouse by taking the difference between X coordinates of the current and the previous sample. The data were coded for direction of movement (toward referent/toward distractor) for each sample based on the object referenced as the treasure's location in the given utterance. The cumulative distance participants had moved the mouse toward either object was calculated by summing over the total distance travelled in each direction up until that time bin (taking into account all previous mouse movements in that direction on that trial). The proportion-of-movement measure for each object was then calculated by taking the distance travelled by the mouse toward the given object divided by the total distance travelled on that trial (regardless of X direction).

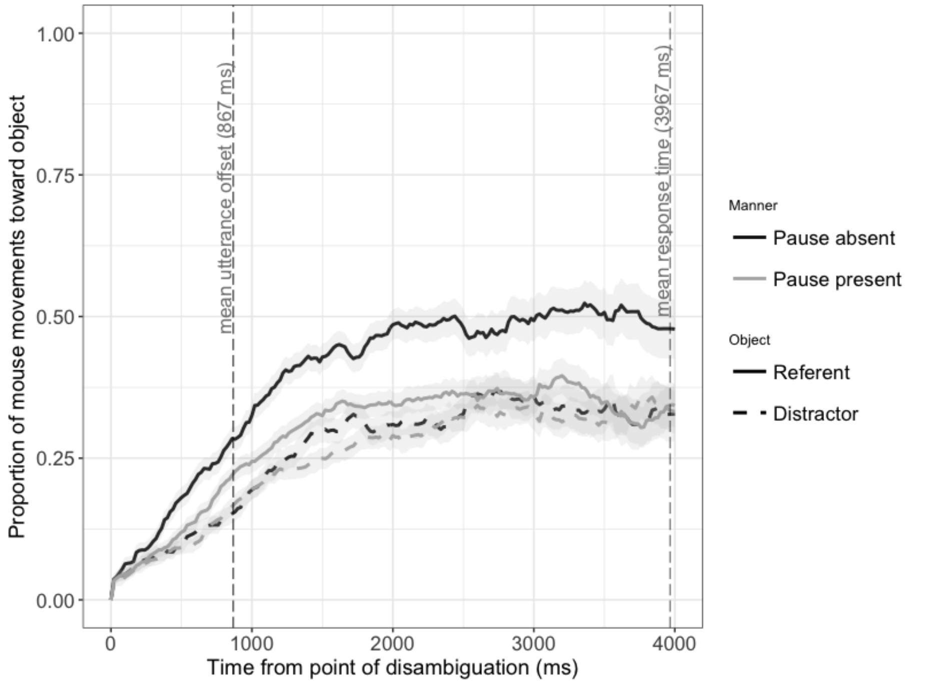
To determine whether the Speaker's manner of delivery influenced Guessers' early inferences about utterance veracity, analyses were conducted over a time window beginning from the point of disambiguation of each utterance to 800 ms post-disambiguation. We focused on Speakers' pauses (identified based on our disfluency categorisation in the main study; see section 6.2.4), since this variable emerged as the strongest predictor of Guessers' veracity judgements. However, to test if there was an effect of other types of disfluency, we included an additional predictor taking into account all other disfluencies (excluding pauses) collapsed into a binary variable. Hence models included the fixed effects of time as well as both forms of manner of delivery (pauses present/absent; other disfluencies present/absent), with time allowed to interact with each of the two forms of manner.[5] All predictors were mean-centred. Models were fitted using empirical logit regression (Barr, 2008), taking as the dependent variable the difference between the e-logit of the proportion of movements toward the referent and toward the distractor. All models included by-subjects and by-items random intercepts and slopes for the three predictors.

Results

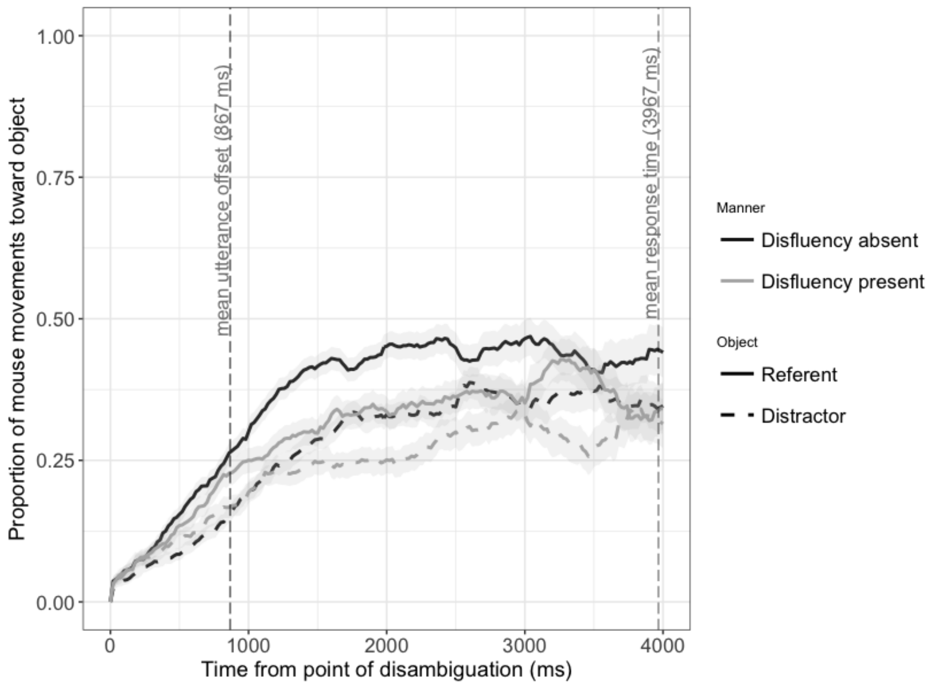
Fig. 6.3 shows the proportion of mouse movements (in terms of distance travelled) toward each object from the point of disambiguation in the utterance to 4000 ms post-disambiguation, based on Speakers' manner of delivery. The average time taken for Guessers to click on an object was 3967 ms post-disambiguation. This time is higher than those observed in our earlier comprehension experiments, and likely reflects the difference between interactive and comprehension-only tasks. The pattern of mouse movements suggests that listeners' early inferences about Speakers' veracity varied with the Speaker's pause behaviour (see Fig. 6.3a). During utterances that did not contain a pause, participants demonstrated a marked preference to move the mouse toward the referent over the distractor that emerged almost immediately post-disambiguation. When the speaker was disfluent in the form of a pause, the same bias was present but noticeably smaller.

Table 6.16 presents the coefficients of the model estimating the effects of time, pauses and other types of disfluency on Guessers' mouse movements. The model revealed a time by manner (pauses) interaction, $\beta = 0.47$, SE $= 0.17$, $t = 2.74$,

[5] The inclusion of the three-way interaction between time and both forms of manner led to model non-convergence, hence this interaction was omitted from the model.



(a)



(b)

Figure 6.3: Mouse-tracking results for Experiment 6.1: Proportion of cumulative distance travelled toward each object (referent or distractor) over time by the speaker's manner of delivery: (a) pauses and (b) other types of disfluency. Shaded areas represent ±1 standard error of the mean.

Table 6.16: Model coefficients for analysis of Guessers' mouse movements.

Model of time, pauses and other types of disfluency on Guessers' mouse movements				
Fixed effects				
		Estimate	SE	t value
(Intercept)		0.31	0.19	1.58
time		0.91	0.37	2.42
manner (pauses)		0.06	0.36	0.18
manner (other disfluencies)		0.14	0.39	0.35
time : manner (pauses)		0.47	0.17	2.74
time : manner (other disfluencies)		0.15	0.16	0.90
Random effects				
		Variance	SD	
participants	(Intercept)	0.61	0.78	
	time	1.80	1.34	
	manner (pauses)	1.11	1.05	
	manner (other disfluencies)	.198	1.42	
items	(Intercept)	1.04	1.02	
	time	5.71	2.34	
	manner (pauses)	7.21	2.69	
	manner (other disfluencies)	6.48	2.55	
No. of obs: 39,509, groups: participant, 24; item, 96				

reflecting the greater referent bias during fluent compared to disfluent utterances characterised by pauses, which increased as the utterance unfolded.

Guessers' mouse movements following utterances characterised by non-pause disfluencies followed a similar trend. Utterances without any non-pause disfluencies exhibited a larger referent bias compared to utterances that did contain non-pause disfluencies (see Fig. 6.3b). However this difference was qualitatively smaller compared to the bias observed in utterances marked by pauses. The model did not yield a time by manner (non-pause disfluency) interaction, $\beta = 0.15$, SE = 0.16, $t = 0.90$, suggesting that Guessers' early inferences about utterance veracity were not affected by other types of disfluencies produced by Speakers.

Discussion

In line with findings from the main study, our mouse-tracking results demonstrate that Guessers' veracity judgements were sensitive to the Speaker's manner of delivery: Utterances characterised by pauses, either filled or silent, were more likely to be judged as lies than those without pauses. More importantly, these effects were observed during our early time window of analysis from 0–800 ms after the point of disambiguation in the utterance. This suggests that Guessers' judgements about utterance veracity evolved rapidly, almost as soon as they could infer a lie based on the Speaker's description, and notably even before Speakers had completed their utterance. Thus, extending findings from our earlier comprehension experiments, we show that these rapid, manner-based judgements take place even in interactive circumstances, where listeners are under conversational and task demands to evaluate speakers' utterances in real-time.

The same bias was not found when we compared utterances containing other (non-pause) disfluencies to those without. This suggests that while listeners may attend to a speaker's pause behaviour to infer utterance veracity, other forms of disfluency may not hold the same perceptual relevance. These results echo an earlier finding by Corley and colleagues that filled and silent pauses have a facilitative effect on listeners' recall abilities (Corley et al., 2007; MacGregor et al., 2010), but repetitions do not (MacGregor, Corley, & Donaldson, 2009). The finding that pauses stand apart from other disfluencies also lends weight to some theories of disfluency production, which propose different functions for different forms of disfluency: Clark's (1996) account, for example, suggests that certain disfluencies, in particular filled pauses, are produced as listener-oriented signals as part of conversation management strategy (cf. Bortfeld et al., 2001). Other forms of disfluency, such as repairs, on the other hand, may be associated with self-monitoring by the parser during planning or articulation (Levelt, 1989), and hence less likely to hold cue potential to a listener.[6]

We emphasise, however, that our mouse-tracking results should be treated with caution. In particular, we note that the analyses presented several difficulties, which may undermine the validity of our conclusions. The first difficulty relates to the nature of the utterances produced by Speakers. The visual similarity between the two objects in each display meant that Speakers had to use complex, often elaborate NPs to distinguish between them. This is reflected in the range of variation we observed in Speakers' descriptions across the experiment. This led to difficulty in establishing the exact point of disambiguation in each utterance. Reasonable estimates were made, based on a contrastive set evoked by Speakers' descriptions on each trial (e.g., closed bin as opposed to the *open* one); however these were determined subjectively post-hoc, and may not reflect Guessers' expectations during the actual game.

The second difficulty stemmed from the interactive nature of the task, which meant that Guessers could ask Speakers for additional information, or attempt various strategies to expose deceptive intent. These trials had to be excluded from analysis, since it was impossible to determine when a reasonable point of disambiguation occurred over the course of the dialogue. The interactive design also meant that some Guessers spent more time deliberating about their judgements, as no time limit was imposed on a response. This is evidenced by the longer time taken for an object click to be detected here compared to our earlier comprehension paradigms. While our results reveal an early effect of Speakers' pauses on Guessers' veracity judgements, the longer response time implies

[6]Note that we are not suggesting that Speakers produced filled pauses as signals of deceptive intent here (in fact, if anything, the data seem to suggest the opposite). However our results support the general idea that there are systematic differences between different forms of disfluency, and that these differences may manifest differently for speakers and listeners.

that Guessers' inferences likely unfolded over a larger time period, suggesting the benefit of a more thorough investigation on the overall time course of listeners' decision-making in such interactive contexts.

Although the dialogic aspect of the task led to some difficulty mapping mouse movements to speech, we maintain that such a design has its benefits, by moving away from the artificiality that characterises many lie perception paradigms. Real-world deception (or any interaction, for that matter) is rarely uni-directional, with listeners passively waiting to form a judgement on speakers' utterances. Rather, it is a dynamic, reciprocal event (Buller & Burgoon, 1996; Burgoon & Buller, 2008; cf. Garrod & Pickering, 2004), with both parties contributing to shape the overall outcome of a discourse. While only tentative conclusions can be drawn from the mouse movements here, they represent a step toward better understanding the unfolding dynamics of interaction in contexts of deception.

6.3 Chapter discussion

This chapter set out to explore the production and perception of manner-based cues to deception. The experiment had two main aims: To test whether a) our main finding from Chapters 4 and 5 that listeners associate disfluency with lying still holds in an ecological context where speakers and listeners interact in real-time, and b) the cues to deception that listeners rely on align with those that speakers actually produce when lying. To that end, we modified the paradigm from the earlier chapters into an interactive game of deception in which speakers and listeners attempted to outwit each other by producing (speakers) and assessing (listeners) a mix of true and false utterances about the treasure's location.

With respect to our first aim, our results demonstrate that listeners still attend to speakers' disfluencies, in particular their pause behaviour, to infer when lying has occurred. This association holds even in the presence of other cues that listeners can potentially draw on, in both the visual and auditory channels of communication. Notably, evidence from the mouse movements suggests that this effect unfolds during the early stages of comprehension, despite competition for attention from other cue behaviours as well as the conversational demands of an interactive task. This provides evidence that our lab-based findings carry over to a more ecologically-valid context where individuals interact face-to-face in a naturalistic setting involving potential deception. Although this deception transpired within a game scenario, we emphasise that this brings us a step closer to real-word deception from earlier paradigms that mainly elicit instructed lies from speakers (e.g., Burgoon & Floyd, 2000; Duran, Dale, & McNamara, 2010) or post-hoc discrimination judgements from listeners (e.g., C. L. Hart et al., 2009; Vrij et al., 2001). Moreover, the dynamics of the gameplay creates an environment in which speakers and listeners are naturally motivated to outwit each other by successfully deceiving and judging deception respectively (cf. Sip et al., 2010). This lends ecological support to the utterances produces by speakers and the inferences made by listeners in the experiment.

With respect to our second goal, we found an unexpected result in a mismatch between the cues that listeners used to index lie behaviour and those that co-occurred with speakers' lies. In particular, although listeners associated speakers' pauses with lying, speakers were actually less likely to pause when lying compared to truth-telling. The finding that listeners still quickly and persistently perceived pauses to index falsehood, even in the face of real-time counter-evidence from speakers, delineates a firm association that is difficult to overcome. This finding provides further insight into listeners' inferential processes when discerning truth from lie both here as well as in our earlier experiments. Recall that in Experiment 5.1, we discussed the possibility that listeners may have considered truth-telling with the intention to mislead more effortful than straightforward lying, leading to a greater expectation that disfluency indexed truth-telling in a speaker perceived as typically truthful (see section 5.2.3). The design of our comprehension paradigm precluded the possibility of ruling out such an explanation. However, the current finding that listeners still associate disfluency with lie behaviour, even when presented with immediate feedback to suggest a misplaced bias, undermines explanations based on such an association between disfluency and truth-telling. Instead, they emphasise the stereotypes of lying that listeners hold, the rapid effect these have on comprehension, and the difficulty in overcoming these biases when inferring truth from lie.

The mismatch between speaker and listener behaviour also suggests that the production and perception of cues to lying may be predicted by different accounts within the framework of deception. Listeners' biases suggest that pauses are interpreted under a cognitive hypothesis, which underscores the cognitive load (and resultant cue behaviours) associated with constructing a lie; speakers' pause behaviour, however, is in line with the attempted control hypothesis, which emphasises that behaviours perceived as potential cues to lying may be suppressed in an effort to maintain a credible front. This disparity highlights the potential pitfalls of extending models of lie production to listeners' comprehension of deception (see section 4.2.10 for a parallel discussion on speaker vs. listener accounts of disfluency). Although listeners demonstrate a tendency to hold persistent notions about stereotypes of deceit (cf. Akehurst et al., 1996; Zuckerman, Koestner, & Driver, 1981), speakers appear able to reason about these biases and manipulate their behaviour accordingly. It remains unclear whether the behavioural control exhibited by our speakers was based on reasoning about their own biases as listeners (i.e. speakers are aware that people generally believe liars are more disfluent) or a direct response to listeners' prejudices within the context of the experiment. In other words, in a different scenario where listeners consistently associate pauses with truth-telling, would speakers modify their behaviour accordingly and produce more pauses when lying? Although this question lies beyond the scope of the current study, it underscores the importance for future research to take into account both speaker and listener perspectives, as well as their respective behaviours within the specific context, to contribute to a more cohesive understanding of deception as the social, communicative phenomenon that it is.

Chapter 7

General discussion and conclusions

The experiments reported in this thesis examined how a speaker's manner of delivery influences a listener's pragmatic inferences, with a particular interest in the time course of these effects during on-line comprehension. We focused on two contexts in which pragmatic inferences arise: The comprehension of scalar expressions and the perception of deception. For both contexts, we explored how these inferences unfolded using an eye- and mouse-tracking paradigm, in which listeners selected between two objects on a display based on a speaker's fluent or disfluent delivery of an utterance. The final experiment situated the comprehension paradigm in an interactive context with the aim of investigating the role of manner of delivery during deception from both a production and a perception perspective. In this chapter, we review our findings in relation to the questions of this thesis and discuss the broader implications of our results.

7.1 The role of manner of delivery

A consistent result that emerged across the six experiments was the effect of the speaker's manner of delivery on listeners' pragmatic inferences about the utterance. In Experiment 3.1, we found that in a context where a literal interpretation of an ambiguous quantifier *some* threatens a speaker's positive self-image, fluent utterances biased listeners to assign a pragmatic meaning of *not all* to *some*, while the presence of disfluency reduced this tendency in favour of a literal (*possibly all*) interpretation. This suggests that listeners interpret a speaker's disfluencies within the social context in which they occur to influence their overall interpretation of the utterance. In the experiments that followed (Chapters 4–6), we extended this finding to contexts where the veracity of the speaker is called into question. Here, we found that fluent utterances biased listeners toward perceiving the speaker to be truthful, while disfluency biased them to infer that the speaker was lying. This bias was found to hold regardless of whether disfluency occurred at the start of the utterance (Experiment 4.1) or in the middle (Experiment 4.2), and in the face of other sources of information that could modulate the effect, such as expectations based on a speaker's likelihood of lying (Experiment 5.1) or the speaker's accent (Experiment 5.2). Notably, this association was found to be robust in a real-life, interactive context, in the immediate presence of multiple competing cues that listeners could draw on in both the visual and auditory modalities (Experiment 6.1). Together, these findings testify to the firm beliefs that perceivers hold about cues to deception, regardless of whether or not these beliefs are founded on actual evidence (cf. Akehurst et al., 1996).

Indeed, although listeners' sensitivity to disfluency in Experiment 6.1 aligned with our earlier comprehension results, their expectations revealed a mismatch with the actual behaviour exhibited by speakers in the experiment. One reason why listeners may hold such potentially misguided beliefs about cues to lying is due to the nature of interactions that involve deception. In a typical, day-to-day deceptive scenario, perceivers often do not receive reliable feedback on the accuracy of their judgements of deception (Ekman & O'sullivan, 1991; Vrij & Semin, 1996). This absence of any confirmatory or contradictory evidence may perpetuate stereotypical associations about deceptive behaviour based on over-simplified rules and heuristics. Over time, these associations may become so ingrained that they are difficult to change (cf. Granhag et al., 2005), even in the face of counter evidence, as our current results reveal.

7.2 Time course of effect

Beyond its effect on listeners' off-line interpretations, this thesis was interested in exploring how manner of delivery influences listeners' inferential processes during real-time comprehension. To that end, we measured listeners' eye and mouse movements in response to unfolding speech. This approach naturally extends sentence verification and response time measures employed by earlier studies by capturing the fine-grained changes that occur during interpretation, affording a clearer picture of the ongoing processes underlying comprehension.

The time course of listeners' eye and mouse movements highlights the speed with which the comprehension system integrates information based on a speaker's manner of delivery to shape listeners' pragmatic inferences. Across six experiments, we found consistent evidence for this effect of manner during a time window corresponding to the earliest stages of comprehension, almost as soon as listeners could assign meaning to the critical linguistic input. These effects were not limited to one aspect of pragmatic comprehension: In Chapter 3 we showed that disfluency influences a listener's on-line interpretation of an ambiguous scalar quantifier; in Chapters 4–6 we showed that this effect extends to listeners' perception of utterance veracity when entertaining the possibility of deception. These results provide converging evidence that listeners rapidly reason about a speaker's disfluencies in contextually flexible ways to guide their unfolding pragmatic inferences during real-time comprehension. Our findings add to a small but growing literature demonstrating how the way a message is conveyed very quickly influences a listener's pragmatic hypotheses about the meaning of an utterance (Kurumada et al., 2014; Van Berkum et al., 2008).

Beyond the effect of the speaker's manner of delivery, our results inform a broader question that this thesis is concerned with regarding the time course of pragmatic interpretation. The speed with which listeners' pragmatic inferences developed in our experiments is incompatible with timing predictions by the standard pragmatic model. This account maintains a two-step approach to pragmatic comprehension, where pragmatic enrichment occurs at a relatively late stage of processing, and only after a context-independent, literal meaning is computed. Contrary to this, our results demonstrate that listeners' pragmatic inferences emerged rapidly, and provide no evidence to suggest a temporal prioritisation of literal understanding. This pattern is broadly consistent with a one-step model of comprehension, in which various sources of contextual information are brought to bear immediately to influence pragmatic meaning from the earliest moments of comprehension.

We note, however, that the experiments in this thesis forefront the perceptual relevance of one such source of information, that of a speaker's disfluency. Indeed, studies that demonstrate support for a one-step model have largely focused on the contribution of a single aspect of contextual information, such as the speaker's co-speech gestures or the listener's knowledge about the world (e.g., Dick et al., 2009; Hagoort et al., 2004; see section 2.3.2). However, understanding language in a typical interpersonal exchange is often situational, requiring listeners to combine streams of information from different domains and modalities (cf. Hagoort, 2005; Taraban & McClelland, 1990). This applies not only to comprehension in cases that involve potential deception, but also to contexts where understanding may be shaped by various sources of information beyond the content of an utterance, such as acquiring new information (Goldin-Meadow & Wagner, 2005), assessing a speaker's certainty (Krahmer & Swerts, 2005), or evaluating a speaker's expertise (Mol et al., 2013). While results across studies currently highlight the immediate relevance of individual constraints on pragmatic interpretation, it remains an open question for future research to explore the time course of the interplay between multiple contextual constraints, and to integrate these within a coherent model of pragmatic comprehension.

7.3 The importance of context

The results of this thesis highlight the influence of manner of delivery on listeners' pragmatic inferences; however, we note that this finding cannot be attributed to its effect as a context-independent cue to understanding. As mentioned earlier, communication and comprehension is highly situational, with meaning often rapidly coloured by the wider context in which an expression occurs (e.g., Nieuwland & Van Berkum, 2006). This is particularly the case for disfluencies, which have been shown to signal various things (e.g., infelicity, uncertainty, speaker difficulty, deception, unpleasant information) depending on the context of the discourse and the listener's comprehension goals (J. E. Arnold et al., 2007; Barr & Seyfeddinipur, 2010; Bonnefon et al., 2015; Bortfeld et al., 2001; Brennan & Williams, 1995; Corley et al., 2007; Zuckerman, Koestner, & Driver, 1981; see Corley & Stewart, 2008 for a review). In other words, listeners interpret disfluencies in flexible ways, taking into account what they may mean given the particular situation and utterance. Accordingly, in Experiment 3.1, the scenario established by our fictitious cover story allowed speakers' filled pauses to function as a signal of a potential face-loss associated with greed (see section 3.2), thereby prompting listeners to assume the more socially dispreferred meaning of *some* (in this case the literal interpretation) within the given context. It is unlikely that in an alternative scenario, where a literal meaning of *some* carried positive rather than negative connotations, a disfluency would result in the same bias in favour of the literal interpretation (cf. Bonnefon et al., 2015, Experiment 2).

In Experiments 4 and 5, we similarly employed cover stories to create a relevant context in which utterances (and accompanying disfluencies) were ostensibly produced by a speaker who was incentivised to deceive listeners by lying some of the time about the treasure's location (see section 4.2.3). This allowed us to situate the experimental task within a game purportedly about lie detection, motivating listeners to respond as they would to authentic deception. Our later finding in Experiment 6.1 that listeners were similarly sensitive to speakers' pauses when performing the same task in an interactive paradigm involving actual deception lends ecological weight to our lab-based findings from Chapters 4 and 5. Collectively, the experiments in this thesis underscore the fact that language comprehension is inherently a context-bound phenomenon. We further emphasise that this should be an important consideration in experimental design: By establishing a relevant domain of interpretation in which listeners can situate what is said, researchers can better understand how the comprehension system depends on the wider context to construct a meaningful message from the speaker's utterance.

7.4 Capabilities and limitations of mouse-tracking

In this thesis, we employed an eye- and mouse-tracking methodology to study how listeners' pragmatic inferences unfold during real-time language comprehension. Besides informing us about the temporal character of pragmatic inferencing, our time course results are additionally relevant to psycholinguistic research from a methodological perspective. Extending studies that have used mouse-tracking to replicate existing eye-tracking paradigms (e.g. Farmer et al., 2007; Spivey et al., 2005), we demonstrate that the two methods can be successfully combined in a single paradigm, yielding temporally and spatially sensitive data on the mechanisms underlying linguistic processing.

Results from our lab-based comprehension paradigms (Chapters 3–5) demonstrate that listeners' mouse movements tend to pattern consistently with their eye movements. Together, the two offer a corroborating account of the ongoing dynamics of comprehension. This finding provides real-time experimental evidence for an intimate link between the linguistic, visual and motor systems: Manual actions such as a mouse movement in response to speech are not merely the end product of comprehension; rather they are part and parcel of cognition, evolving as the comprehension system integrates linguistic input with the co-present visual information. This opens up opportunities for researchers wishing to study populations which may present challenges in eye-tracking, such as certain clinical or developmental groups (cf. Sasson & Elison, 2012). The portability of mouse-tracking also offers promise as a tool for studies seeking to investigate populations that may be difficult to access, or to obtain time course data on a large-scale, through employing mouse-tracking in web-based data collection. Future research may further exploit this relationship between action and cognition by tracking other forms of manual gestures, such as language-mediated hand or arm movements (e.g. Duran et al., 2010), to enrich our knowledge of how multiple processes may interact to constrain the manual and mental trajectories of comprehension.

As with any methodology, however, mouse-tracking is not without its limitations. Unlike stimulus-driven eye movements which are often reflexive (Theeuwes, Kramer, Hahn, & Irwin, 1998), the goal-oriented nature of mouse movements means that they may not be executed prior to an initial commitment to a response. Experiment 6.1 demonstrates how this may be a potential disadvantage for interactive studies seeking to utilise mouse-tracking, due to the relative lack of control over the latency period to listeners initiating a mouse movement, as well as their time taken to click on a response. This is evident from the mean response time exhibited by listeners in Experiment 6.1, which was almost twice as long as that observed in our earlier comprehension-only paradigms. This naturally complicates the task of establishing an appropriate time window for analysis due to greater variation in the time course of listeners' interpretations. While comprehension-only paradigms can overcome this problem by imposing a time limit on response, the dialogic nature of the task in Experiment 6.1 precluded this possibility, since listeners' mouse movements were contingent on multiple uncontrollable factors such as speakers' initiations of an utterance, the durations of their utterances, and any subsequent dialogue between participants.

As Spivey et al. (2005) note, eye-tracking and mouse-tracking have their respective strengths and weaknesses. With advances in oculomotor research, tracking participants' eye movements is becoming an increasingly portable methodology (Bulling & Gellersen, 2010; Kunze, Utsumi, Shiga, Kise, & Bulling, 2013), providing researchers with the option of obtaining fine-grained time course data in mobile, day-to-day settings. This opens up possibilities for future studies to combine the two measures in interactive paradigms, allowing us to obtain a clearer picture of real-time language use in context.

7.5 Conclusions

How do listeners understand language? The simplest assumption might be that they derive meaning based on the content of utterances. However, despite the fact that listeners are adept at decoding the meaning of words within sentences, this information is often insufficient for fully understanding what a speaker means. In this thesis, we showed how the manner in which an utterance is conveyed influences the meaning derived by listeners. Crucially, this information is extracted and used very quickly by listeners to infer a contextually relevant interpretation during real-time comprehension. Our results challenge the traditional view held by the standard pragmatic model that pragmatic understanding is slow and resource-demanding. Instead, they support a model which integrates pragmatic constraints from the earliest moments of comprehension. The construction of meaning is therefore not merely a process of combining individual word meanings in a context-free, literal analysis; rather, it takes into rapid account the pragmatics of the exchange to shape meaning right from the start.

References

Akaike, H. (1973). Information theory and an extention of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281).

Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 217–235.

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79–B87.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Anolli, L., & Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior*, *21*(4), 259–284.

Appelt, D. E. (1985). Planning English referring expressions. *Artificial Intelligence*, *26*(1), 1–33.

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.

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), 914–930.

Arnold, J. E., & Tanenhaus, M. K. (2011). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. Pearlmutter (Eds.), (pp. 197–217). Cambridge, MA: MIT Press.

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*(9), 578–582.

Arnold, T. W. (2010). Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management*, *74*(6), 1175–1178.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456.

Bailey, K. G., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 485–500). Oxford: Elsevier.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In S. Santi, I. Guaïtella, C. Cave, & G. Konopczynski (Eds.), *Oralité et gestualité, communication multimodale, intéraction* (pp. 597–600). Paris: L'Harmattan.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, *25*(4), 441–455.

Barres, P., & Johnson-Laird, P. N. (2003). On imagining what is true (and what is false). *Thinking & Reasoning*, *9*(1), 1–42.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=lme4` (R package version 1.1-6)

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.

Beňuš, S. (2009). Variability and stability in collaborative dialogues: Turn-taking and filled pauses. In *Proceedings of the Tenth Annual Conference of the International Speech Communication Association* (pp. 796–799). Brighton, UK.

Beňuš, S., Enos, F., Hirschberg, J. B., & Shriberg, E. (2006). Pauses in deceptive speech. In *Proceedings of the ISCA 3rd International Conference on Speech Prosody.* Dresden, Germany.

Blank, G. D. (1988). Metaphors in the lexicon. *Metaphor and Symbolic Activity*, *3*(1), 21–36.

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer software manual]. Retrieved from `http://www.praat.org`

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

Bond, C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior*, *14*(3), 189–204.

Bond, M. H., & Lai, T.-M. (1986). Embarrassment and code-switching into a second language. *Journal of Social Psychology*, *126*(2), 179–186.

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.

Bonnefon, J.-F., De Neys, W., & Feeney, A. (2011). Processing scalar inferences in face-threatening contexts. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3389–3394). Austin, TX: Cognitive Science Society.

Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, *112*(2), 249–258.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, *8*, 148–158.

Boomer, D. S., & Dittmann, A. T. (1964). Speech rate, filled pause, and body movement in interviews. *The Journal of Nervous and Mental Disease*, *139*(4), 324–327.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (2002). Modality independence of word comprehension. *Human Brain Mapping*, *16*(4), 251–261.

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.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, *75*, 104–116.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, *66*(1), 123–142.

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.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity, and intergroup relations* (pp. 119–36). London: Academic Press.

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.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.

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.

Broen, P. A., & Siegel, G. M. (1972). Variations in normal speech disfluencies. *Language and Speech*, *15*(3), 219–231.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge: Cambridge University Press.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242.

Buller, D. B., Comstock, J., Aune, R. K., & Strzyzewski, K. D. (1989). The effect of probing on deceivers and truthtellers. *Journal of Nonverbal Behavior*, *13*(3), 155–170.

Bulling, A., & Gellersen, H. (2010). Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing*, *9*(4), 8–12.

Burgoon, J. K., & Buller, D. B. (2008). Interpersonal deception theory. In L. A. Baxter & D. O. Braithewaite (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 227–239). Thousand Oaks, CA: SAGE Publications.

Burgoon, J. K., Buller, D. B., Floyd, K., & Grandpre, J. (1996). Deceptive realities: Sender, receiver, and observer perspectives in deceptive conversations. *Communication Research*, *23*(6), 724–748.

Burgoon, J. K., & Floyd, K. (2000). Testing for the motivation impairment effect during deceptive and truthful interaction. *Western Journal of Communication (includes Communication Reports)*, *64*(3), 243–267.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35.

Busso, C., & Narayanan, S. S. (2007). Interrelation between speech and facial gestures in emotional utterances: A single subject study. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(8), 2331–2347.

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*(3), 193–204.

Cantoni, E., & Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, *96*(455), 1022–1030.

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.

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.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond* (Vol. 3, pp. 39–103). Oxford: Oxford University Press.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Christenfeld, N. (1994). Options and ums. *Journal of Language and Social Psychology*, *13*(2), 192–199.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H. (2002). Speaking in time. *Speech Communication*, *36*(1), 5–13.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguists*. New York: Harcourt Brace Jovanovich.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experiment Psychology: General*, *113*(1), 121–126.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–372). Amsterdam: Elsevier.

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.

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PloS one*, *6*(5), e19792. doi: https://doi.org/10.1371/journal.pone.0019792

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.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, *2*(4), 589–602.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, *111*(6), 2862–2873.

Cummins, C., & Rohde, H. (2015). Evoking context with contrastive stress: Effects on pragmatic enrichment. *Frontiers in Psychology*, *6*(1779), 1–11.

Curran-Everett, D. (2000). Multiple comparisons: Philosophies and illustrations. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *279*(1), R1–R8.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459.

Da Silva, C. S., & Leach, A.-M. (2013). Detecting deception in second-language speakers. *Legal and Criminological Psychology*, *18*(1), 115–127.

Davydova, J. (2011). *The present perfect in non-native Englishes: A corpus-based study of variation*. Berlin: Walter de Gruyter.

Degen, J. (2015). Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods. *Semantics and Pragmatics*, *8*(11), 1–55.

Degen, J., Reeder, P. A., Carbary, K., & Tanenhaus, M. K. (2009). Using a novel experimental paradigm to investigate the processing of scalar implicatures. In *Paper presented at the 3rd Biennial Conference of Experimental Pragmatics* (pp. 23–25). Lyon, France.

Degen, J., & Tanenhaus, M. K. (2010). Naturalness of lexical alternatives influences interpretation of 'some'. In *Poster presented at the 23rd Annual CUNY Conference on Human Sentence Processing.* New York.

Degen, J., & Tanenhaus, M. K. (2011). Making inferences: The case of scalar implicature processing. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 3299–3304). Austin, TX.

Degen, J., & Tanenhaus, M. K. (2014). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710.

de Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, *19*(2), 85–114.

Dennison, H. Y., & Schafer, A. J. (2010). Online construction of implicature through contrastive prosody. In *Proceedings of the 5th International Conference of Speech Prosody* (pp. 1–4). Chicago, IL.

DePaulo, B. M. (1988). Nonverbal aspects of deception. *Journal of Nonverbal Behavior*, *12*(3), 153–161.

DePaulo, B. M., Blank, A. L., Swaim, G. W., & Hairfield, J. G. (1992). Expressiveness and expressive control. *Personality and Social Psychology Bulletin*, *18*(3), 276–285.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63–79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yullie (Ed.), *Credibility assessment* (pp. 51–70). Dordrecht: Kluwer Academic Publishers.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior*, *12*(3), 177–202.

DePaulo, B. M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. *Journal of Personality and Social Psychology*, *45*(5), 1096–1103.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118.

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.

Desroches, A. S., Joanisse, M. F., & Robertson, E. K. (2006). Specific phonological impairments in dyslexia revealed by eyetracking. *Cognition*, *100*(3), B32–B42.

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, *25*(2-3), 204–222.

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics*, *31*(12), 1579–1599.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, *30*(11), 3509–3526.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental Psychology*, *38*(3), 438–445.

Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, *33*(4), 727–733.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446.

Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin & Review*, *17*(4), 486–491.

Dynel, M. (2011). A web of deceit: A neo-Gricean view on types of verbal deception. *International Review of Pragmatics*, *3*(2), 139–167.

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.

Ekman, P. (1992). *Telling lies: Cues to deceit in the marketplace, politics and marriage.* New York: Norton.

Ekman, P. (2001). *Telling lies: Cues to deceit in the marketplace, politics and marriage* (3rd ed.). New York: Norton.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*(1), 49–98.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of Personality and Social Psychology*, *29*(3), 288–298.

Ekman, P., & O'sullivan, M. (1991). Who can catch a liar? *American Psychologist*, *46*(9), 913–920.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of Nonverbal Behavior*, *15*(2), 125–135.

Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B., & D'Imperio, M. (2016). Intonation in the processing of contrast meaning in French: An eye-tracking study. In *Proceedings of the 6th International Conference of Speech Prosody* (pp. 1225–1229). Boston, MA.

Evans, J. R., & Michael, S. W. (2014). Detecting deception in non-native English speakers. *Applied Cognitive Psychology*, *28*(2), 226–237.

Farabee, D. J., Ramsey, S. L., & Cole, S. G. (1993). Social anxiety and speaker gaze in a persuasive atmosphere. *Journal of Research in Personality*, *27*(4), 365–376.

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.

Feeney, A., & Bonnefon, J.-F. (2012). Politeness and honesty contribute additively to the interpretation of scalar expressions. *Journal of Language and Social Psychology*, *32*(2), 181–190.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*(2), 121–132.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209–221.

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*(5), 921–928.

Fox, D. (2014). Cancelling the Maxim of Quantity: Another challenge for a Gricean theory of scalar implicatures. *Semantics and Pragmatics*, *7*, 1–20.

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.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, *27*(1), 35–53.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, *29*(2), 320–326.

Fox Tree, J. E. (2007). Folk notions of um and uh, you know, and like. *Text & Talk*, *27*(3), 297–314.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, *62*(2), 151–167.

Franke, M., De Jager, T., & Van Rooij, R. (2009). Relevance in cooperation and conflict. *Journal of Logic and Computation*, *22*(1), 23–54.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*.

Frith, C. (2000). The role of dorsolateral prefrontal cortex in the selection of oction as revealed by functional imaging. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes* (pp. 549–565). Cambridge, MA: MIT Press.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form.* New York: Academic Press.

Gerring, R. J., & Healy, A. F. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 667.

Geurts, B. (2010). *Quantity implicatures.* Cambridge: Cambridge University Press.

Gibbs, R. W. (1986a). Comprehension and memory for nonliteral utterances: The problem of sarcastic indirect requests. *Acta Psychologica*, *62*(1), 41–57.

Gibbs, R. W. (1986b). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*(1), 3–15.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge: Cambridge University Press.

Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*(4), 457–486.

Gibbs, R. W., & Colston, H. (2006). Figurative language. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 835–862). London: Academic Press.

Gibbs, R. W., & Gerrig, R. J. (1989). How context makes metaphor comprehension seem 'special'. *Metaphor and Symbol*, *4*(3), 145–158.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703–712.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 577–590.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association*, *10*(1), 13–48.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206.

Giora, R. (2003). *On our mind: Salience, context, and figurative language.* Oxford: Oxford University Press.

Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol*, *14*(4), 241–257.

Glucksberg, S. (2001). Metaphor: The central trope. In S. Glucksberg & M. S. McGlone (Eds.), *Understanding figurative language: From metaphor to idioms.* Oxford: Oxford University Press.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences*, *7*(2), 92–96.

Glucksberg, S., Gildea, P., & Bookin, H. B. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 85–98.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*(2), 214–237.

Goffman, E. (1967). *Interaction ritual: Essays in face to face behavior.* Garden City, NY: Anchor Books.

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

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234–241.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* Academic Press.

Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs*, *132*(3), 197–214.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184.

Granhag, P. A., & Strömwall, L. A. (2002). Repeated interrogations: Verbal and non-verbal cues to deception. *Applied Cognitive Psychology*, *16*(3), 243–257.

Granhag, P. A., Strömwall, L. A., & Hartwig, M. (2005). Granting asylum or not? Migration board personnel's beliefs about deception. *Journal of Ethnic and Migration Studies*, *31*(1), 29–50.

Gregory, M. E., & Mergler, N. L. (1990). Metaphor comprehension: In search of literal truth, possible sense, and metaphoricity. *Metaphor and Symbol*, *5*(3), 151–173.

Grice, H. P. (1975). Logic and conversation. In P. Coleman & J. L. Morgan (Eds.), *Syntax and semantics volume 3* (pp. 41–58). New York: Academic Press.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Grodner, D., 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.

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In E. A. Gibson & N. J. Pearlmutter (Eds.), *The processing and acquisition of reference* (pp. 239–272). Cambridge, MA: MIT Press.

Gupta, S., Sakamoto, K., & Ortony, A. (2012). Telling it like it isn't—a comprehensive approach to analyzing verbal deception. In F. Paglieri, L. Tummolini, R. Falcone, & M. Miccli (Eds.), *The goals of cognition. Festschrift for Cristiano Castelfranchi* (pp. 579–612). London: London College Publications.

Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423.

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.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*(1481), 801–811.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development*, *62*(1), 83–97.

Hamblin, J. L., & Gibbs, R. W. (2003). Processing the meanings of what speakers say and implicate. *Discourse Processes*, *35*(1), 59–80.

Hancock, J. T., Woodworth, M. T., & Goorha, S. (2010). See no evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation*, *19*(4), 327–343.

Handl, S. (2011). *The conventionality of figurative language: A usage-based study* (Vol. 46). Tübingen: Books on Demand.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105–115.

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.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Happé, F. G. (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. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154.

Hart, C. L., Fillmore, D. G., & Griffith, J. D. (2009). Indirect detection of deception: Looking for change. *Current Research in Social Psychology*, *14*(9), 134–142.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*(3), 169–177.

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*(4), 643–659.

Heyd, T. (2013). Email hoaxes. In S. Herring, D. Stein, & T. Virtanen (Eds.), *Pragmatics of computer-mediated communication* (Vol. 9, pp. 387–410). Berlin: Walter de Gruyter.

Heydon, G. (2008). The art of deception: Myths about lie detection in written confessions. *Het analyseren van de geloofwaardigheid van verhoren. Het gebruik van leugendetectiemethoden*, 171–180.

Honeck, R. P., Welge, J., & Temple, J. G. (1998). The symmetry control in tests of the standard pragmatic model: The case of proverb comprehension. *Metaphor and Symbol*, *13*(4), 257–273.

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

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, form, and use in context: Linguistic applications*, 11–42.

Horn, L. R. (2004). Implicature. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 3–28). Oxford: Blackwell Publishing.

Huang, Y. T., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Huang, Y. T., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739.

Huang, Y. T., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*(8), 1161–1172.

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.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.

Inhoff, A. W., Lima, S. D., & Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Memory & Cognition*, *12*(6), 558–567.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, *2*(3), 4.

Janus, R. A., & Bever, T. G. (1985). Processing of metaphoric language: An investigation of the three-stage model of metaphor comprehension. *Journal of Psycholinguistic Research*, *14*(5), 473–487.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Johnson, R., Henkell, H., Simon, E., & Zhu, J. (2008). The self in conflict: The role of executive processes during truthful and deceptive responses about attitudes. *Neuroimage*, *39*(1), 469–482.

Jucker, A. H., Smith, S. W., & Lüdge, T. (2003). Interactive aspects of vagueness in conversation. *Journal of Pragmatics*, *35*(12), 1737–1769.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Katsos, N., Breheny, R., & Williams, J. (2005). The interaction of structural and contextual constraints during the on-line generation of scalar inferences. In B. G. Bara, L. Barsalou, & M. Bucciarell (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1108–1113). Stresa, Italy.

Katz, A. N., & Pexman, P. M. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. *Metaphor and Symbol*, *12*(1), 19–41.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*(1), 253–260.

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, *28*(4), 375–385.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225.

King, J. P., Loy, J. E., & Corley, M. (2017). Contextual effects on online pragmatic inferences of deception. *Discourse Processes*. doi: 10.1080/0163853X.2017.1330041

Kirk, E., Pine, K. J., & Ryder, N. (2011). I hear what you say but I see what you mean: The role of gestures in children's pragmatic comprehension. *Language and Cognitive Processes*, *26*(2), 149–170.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194.

Kjellmer, G. (2003). Hesitation. in defence of er and erm. *English Studies*, *84*(2), 170–198.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1), 205–223.

Knapp, M. L., & Comaden, M. E. (1979). Telling it like it isn't: A review of theory and research on deceptive communications. *Human Communication Research*, *5*(3), 270–285.

Knapp, M. L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. *Human Communication Research*, *1*(1), 15–29.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*(1), 29–53.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Kunze, K., Utsumi, Y., Shiga, Y., Kise, K., & Bulling, A. (2013). I know what you are reading: Recognition of document types using mobile eye tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers* (pp. 113–116).

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.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 647–652).

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP).

Annual Review of Psychology, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology Supplement*, *39*, 325–330.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328.

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing - evidence from event-related brain potentials. *Neuroscience Letters*, *339*(3), 191–194.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18.

Lehiste, I. (1996). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–239). New York: Academic Press.

Levelt, W. J. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT press.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communications Monographs*, *66*(2), 125–144.

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

Lickley, R. J. (1995). Missing disfluencies. In *Proceedings of the International Congress of Phonetic Sciences* (Vol. 4, pp. 192–195). Stockhom, Sweden.

Lickley, R. J., & Bard, E. G. (1996). On not recognizing disfluencies in dialogue.

In *Proceedings of the 4th International Conference on Spoken Language* (Vol. 3, pp. 1876–1879). Philadelphia, PA.

Lindsay, J., & O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, *24*(2), 101–115.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426.

Loy, J. E., Rohde, H., & Corley, M. (2016a). Effects of disfluency in online interpretation of deception. *Cognitive Science*. doi: 10.1111/cogs.12378

Loy, J. E., Rohde, H., & Corley, M. (2016b, May–June). Lying, in a manner of speaking. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 8*. Boston, MA.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2009). Not all disfluencies are are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, *111*(1), 36–45.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982–3992.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44.

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., . . . Anna di Palma, M. (2016). robustbase: Basic robust statistics [Computer software manual]. Retrieved from `http://robustbase.r-forge.r-project.org/` (R package version 0.92-7)

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*(3), 245–297.

Mann, S., Vrij, A., & Bull, R. (2002). Suspects, lies, and videotape: An analysis of authentic high-stake liars. *Law and Human Behavior*, *26*(3), 365–376.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*(1), 137.

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.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24.

Merlo, S., & Mansur, L. L. (2004). Descriptive discourse: Topic familiarity and disfluencies. *Journal of Communicative Disorders*, *37*(6), 489–503.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition*, *108*(1), 271–280.

Mol, L., Kuhlen, A., van der Steen, R., & Obbens, M. (2013). Beliefs about a speaker affect feeling of another's knowing. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society.* Austin, TX.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306.

Murphy, G. L. (1985). Processes of understanding anaphora. *Journal of Memory and Language*, *24*(3), 290–303.

Murtaugh, P. A. (2014). In defense of P values. *Ecology*, *95*(3), 611–617.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *The Journal of the Acoustical Society of America*, *82*(5), 1512–1524.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, *85*(2), 203–210.

O'Connell, D. C., & Kowal, S. (2004). The history of research on the filled pause as evidence of The Written Language Bias in Linguistics (Linell, 1982). *Journal of Psycholinguistic Research*, *33*(6), 459–474.

Ortony, A., Reynolds, R. E., & Arter, J. A. (1978). Metaphor: Theoretical and empirical research. *Psychological Bulletin*, *85*(5), 919–943.

O'Sullivan, M., Ekman, P., & Friesen, W. V. (1988). The effect of comparisons on detecting deceit. *Journal of Nonverbal Behavior*, *12*(3), 203–215.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech & Language*, *9*(1), 19–35.

O'Connell, D. C., & Kowal, S. (2005). Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research*, *34*(6), 555–576.

Pagano, A. M., & Arnold, T. W. (2009). Detection probabilities for ground-based breeding waterfowl surveys. *Journal of Wildlife Management*, *73*, 392–398.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–282.

Pfaff, K. L., Gibbs, R. W., & Johnson, M. D. (1997). Metaphor in using and understanding euphemism and dysphemism. *Applied Psycholinguistics*, *18*(1), 59–83.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*(4), 434–446.

Pon-Barry, H., & Shieber, S. M. (2011). Recognizing uncertainty in speech. *Eurasip Journal on Advances in Signal Processing*.

Pynte, J., Besson, M., Robichon, F. H., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain and Language*, *55*(3), 293–316.

Quek, F., McNeill, D., Bryll, R., Kirbas, C., Arslan, H., McCullough, K. E., … Ansari, R. (2000). Gesture, speech, and gaze cues for discourse segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Vol. 2, pp. 247–254).

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: SAGE publications.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*(3), 337–353.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71–78). New York.

Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. *JoVE (Journal of Visualized Experiments)*(61), e3675–e3675.

Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*(3), 362–367.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 750–755). Vancouver, BC.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schwoebel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: Further evidence. *Metaphor and Symbol*, *15*(1-2), 47–61.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language.* Cambridge: Cambridge University Press.

Searle, J. R. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 92–123). New York: Cambridge University Press.

Sedivy, J. C. (2001). Evidence of Gricean expectations in on-line referential processing. In *Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing.* Philadelphia, PA.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York:

Routledge.

Shriberg, E. (1996). Disfluencies in switchboard. In *Proceedings of International Conference on Spoken Language Processing* (Vol. 96, pp. 11–14). Philadelphia, PA.

Shriberg, E. (2001). To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, *31*(1), 153–169.

Siegel, G. M., Lenske, J., & Broen, P. (1969). Suppression of normal speech disfluencies through response cost. *Journal of Applied Behavior Analysis*, *2*(4), 265–276.

Siegman, A. W., & Pope, B. (1966). Ambiguity and verbal fluency in the TAT. *Journal of Consulting Psychology*, *30*(3), 239.

Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., & Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia*, *48*(12), 3619–3626.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*(1), 160–164.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38.

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130.

Snipes, M., & Taylor, D. C. (2014). Model selection and Akaike Information Criteria: An example from wine ratings and prices. *Wine Economics and Policy*, *3*(1), 3–9.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human, Learning, and Memory*, *6*(2), 174–215.

Spence, S. A., Kaylor-Hughes, C., Farrow, T. F., & Wilkinson, I. D. (2008). Speaking of secrets and lies: The contribution of ventrolateral prefrontal cortex to vocal deception. *Neuroimage*, *40*(3), 1411–1418.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398.

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.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, *20*(4), 421–446.

Stanley, G., & Martin, D. S. (1968). Eye-contact and the recall of material involving competitive and noncompetitive associations. *Psychonomic Science*, *13*(6), 337–338.

Staum Casasanto, L. (2008). Does social information influence sentence processing? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 799–804). Washington, DC.

Staum Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 127–132).

Storto, G., & Tanenhaus, M. K. (2005). Are scalar implicatures computed online? In E. Maier, C. Bary, & J. Huitink (Eds.), *Proceedings of Sinn und Bedeutung 9* (pp. 431–445). Nijmegen.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, *18*(1), 86–100.

Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, *12*(2), 207–219.

Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *The Economic Journal*, *119*(534), 47–60.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*, 485–496.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94.

Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, *43*(3), 804–810.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 309–26). Amsterdam, NL: Elsevier.

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*, 1632–1634.

Taraban, R., & McClelland, J. L. (1990). Parsing and comprehension: A multiple-constraint view. In D. A. Balota, G. G. F. d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 231–263). Hillsdale, NJ: Erlbaum.

Tavakoli, P. (2010). Pausing patterns: Differences between L2 learners and native speakers. *ELT journal*, *65*(1), 71–79.

Temple, J. G., & Honeck, R. P. (1999). Proverb comprehension: The primacy of literal meaning. *Journal of Psycholinguistic Research*, *28*(1), 41–70.

ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychological Science*, *25*(5), 1098–1105.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385.

Tomlinson, 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.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(4), 558–592.

Tsuchiya, M., Kogure, S., Nishizaki, H., Ohta, K., & Nakagawa, S. (2008). Developing corpus of Japanese classroom lecture speech contents. In N. Calzolari et al. (Eds.), *Proceedings of the 6th International Conference on Language Resources and Education.* Marrakech, Morocco: European Language Resources Association (ELRA).

Turner, R. E., Edgley, C., & Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology*, 69–89.

Van Berkum, J. J. (2012). The electrophysiology of discourse and conversation. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 589–614). New York: Cambridge University Press.

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.

Van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*(3), 701–718.

Van Tiel, B., Van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar diversity. *Journal of Semantics*, *33*(1), 137–175.

Villar, G., Arciuli, J., & Mallard, D. (2012). Use of "um" in the deceptive speech of a convicted murderer. *Applied Psycholinguistics*, *33*(01), 83–95.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of Social Psychology*, *133*(5), 601–610.

Vrij, A. (1995). Behavioral correlates of deception in a simulated police interview. *The Journal of Psychology*, *129*(1), 15–28.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Chichester: Wiley.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology*, *11*(2), 297–312.

Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence*, *7*(1), 25–36.

Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin*, *27*(7), 899–909.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*(4), 239–263.

Vrij, A., Ennis, E., Farman, S., & Mann, S. (2010). People's perceptions of their truthful and deceptive interactions in daily life. *Open Access Journal of Forensic Psychology*, *2*, 6–49.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110–117.

Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law*, *5*(3), 203–215.

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology*, *15*(2), 187–203.

Vrij, A., Mann, S., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, *32*(3), 253–265.

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, *20*(1), 65–80.

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior*, *15*(3), 169–184.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905–945.

Wardlow Lane, L., & Liersch, M. J. (2012). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes*, *27*(3), 462–473.

Watanabe, M. (2002). Fillers as indicators of discourse segment boundaries in Japanese monologues. In *Proceedings of the 1st International Conference on Speech Prosody* (pp. 691–694). Aix-en-Provence, France: Laboratoire Parole et Langage.

Wenzel, A., Graff-Dolezal, J., Macho, M., & Brendle, J. R. (2005). Communication and social skills in socially anxious and nonanxious individuals in the context of romantic relationships. *Behaviour Research and Therapy*, *43*(4), 505–519.

White, S. (1989). Backchannels across cultures: A study of Americans and Japanese. *Language in Society*, *18*(1), 59–76.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation.

Journal of Memory and Language, *31*(2), 183–194.

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235–1249.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation.* Genoa.

Womack, K., McCoy, W., Alm, C. O., Calvelli, C., Pelz, J. B., Shi, P., & Haake, A. (2012). Disfluencies as extra-propositional indicators of cognitive processing. In *Proceedings of the Workshop on Extra-propositional Aspects of Meaning in Computational Linguistics* (pp. 1–9).

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods*, *7*(5), 397–429.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–59). San Diego, CA: Academic Press.

Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*(1), 105–114.