# Coordinating utterances during conversational dialogue: The role of content and timing predictions

Ruth Corps

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

to

Department of Psychology
School of Philosophy, Psychology, and Language Sciences
The University of Edinburgh

2018

# Declaration

I hereby declare:

(a) that this thesis is of my own composition, and

(b) that the work reported in this thesis has been carried out by myself, except where acknowledgement of the work of others is made in text, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified in text, and

(d) that the included publications are my own composition.

The following chapters of this thesis based on manuscripts that have been published or submitted to peer-reviewed journals:

Chapter 1: Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes, 55,* 230-240. Authorship details: Corps wrote the original manuscript and Gambi and Pickering acted as supervisors and contributed to the revision of the manuscript.

Chapter 2: Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition, 175,* 77-95. Authorship details: Corps designed the study, ran the participants, analyzed the data, and wrote the original manuscript. Crossley conducted data collection for Experiment 1. Gambi and Pickering acted as supervisors, gave feedback on each of these steps, and contributed to the revision of the manuscript.

Chapter 3: Corps, R. E., Gambi, C., & Pickering, M. J. (under review). How do listeners time response articulation during conversational turn-taking? The role of speech rate entrainment. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Authorship details: Corps designed the study, ran the participants, analyzed the data, and wrote the original manuscript. Gambi and Pickering acted as supervisors, gave feedback on each of these steps, and contributed to the revision of the manuscript.

Ruth Corps

# Abstract

During conversation, we take turns at talk and switch between listening to a speaker and producing an appropriate and timely response. In fact, we often do so with relatively little gap or overlap between our own and our partner's contribution. Some theories argue that we manage this process by predicting what we are going to hear. For example, if a speaker says *I would like to go outside to fly a…*, then the listener may predict that the speaker's next word will likely be *kite*. However, little is known about how these predictions aid coordination during conversational dialogue. In particular, how does prediction help listeners comprehend the speaker's turn, prepare a response (i.e., decide what they want to say), and time its articulation (i.e., decide when they want to say it)? And to what extent are these processes interwoven? This thesis firstly addressed this issue by presenting participants with questions in which they either could (e.g., *Are dogs your favorite animal?*) or could not (e.g., *Would you like to go to the supermarket*?) predict the speaker's final word. We asked them to either complete a button-pressing task (Experiments 1 and 3), in which they indicated when they thought the speaker would reach the end of their utterance, or a question-answering task (Experiment 2 and 4), in which they verbally answered each question either *yes* or *no*. We found that listeners responded earlier in the question-answering task when the final word(s) of the question were predictable rather than unpredictable. However, we found no effects of content or length predictability on the precision (i.e., how closely participants responded to the speaker's turn-end) of participants' button-presses or verbal responses. Thus, the results of Experiments 1-4 suggest that listeners use content predictions to prepare a response, but not to predict turn-endings. In other words, preparation and articulation relied on different mechanisms. Experiments 5 and 6 also used a question-answering task and provided further support for this conclusion. In particular, we manipulated the speech rate of the context (e.g., *Do you have a…*) and the final word (e.g., *dog?*) of questions using time-compression, so that each component was spoken at the natural rate or twice as fast. We found that participants responded earlier when context was speeded rather than natural, suggesting they entrained to the speaker's context rate, which in turn influenced when they launched articulation. We also found that listeners responded earlier when the speaker's final word (consisting of a single syllable) was speeded rather than natural, regardless of context rate, suggesting they updated their entrainment after encountering a single syllable at a different rate. In Experiment 6, this final word effect occurred regardless of whether the speaker's final word was predictable or unpredictable, suggesting that speech rate entrainment was used to time articulation independently from preparing the content of a response. Finally, since response preparation and timing articulation rests on successfully comprehending the speaker's turn, Experiments 7-9 investigated how prediction helps listeners understand distorted speech by presenting participants with question-answer sequences, in which the answer was distorted.  Results suggested that comprehension of the distorted answer was sensitive to the plausibility of the answer, rather than the predictability of the question, suggesting that understanding distorted speech is driven by ease of integration but not prediction. Together, these studies provide insight into the role that prediction plays in comprehension, response preparation, and articulation.

# Lay Summary

During conversation, we take turns at talk and switch between listening to a speaker and producing an appropriate and timely response. In fact, we often do so with relatively little gap or overlap between our own and our partner's contribution. Some theories argue that we manage this process by predicting what we are going to hear. For example, if a speaker says *I would like to go outside to fly a…,* then the listener may predict that the speaker's next word will likely be *kite*. In this thesis, I investigate how these predictions aid coordination during conversational dialogue by testing their role in three different, but related, mechanisms: (1) Preparing the content of a response, (2) timing its production, so that there is little overlap or gap between turns, and (3) understanding the speaker in difficult circumstances, such as when their utterances are distorted. First, I used a *yes/no* question-answering task to investigate how listeners use predictions of what the speaker is going to say (i.e., the content of the speaker's utterance) to decide how they themselves wish to respond. From this, I was able to determine how far in advance listeners prepare their own response. To determine how listeners time production of this response, I manipulated the speech rate of utterances. I also conducted additional experiments in which participants pressed a button when they expected the speaker to reach the end of their utterance. Finally, I investigated the role of prediction in understanding speech in difficult circumstances by presenting participants with distorted speech under conditions in which they either did or did not know what the speaker was going to say.

# Acknowledgements

I would not have been able to complete this thesis without the invaluable help of numerous people. First, I would like to thank my supervisors, Prof. Martin Pickering, Dr. Chiara Gambi, and Dr. Hugh Rabagliati, for their support, encouragement, and guidance throughout the last three years – your suggestions have shaped the research in this thesis and have helped me develop my research career. Martin, thank you for useful theoretical discussions, for always being enthusiastic about my results, and for offering advice on academic issues in general (and for agreeing to work with me post-PhD!). Chiara, thank you for teaching me mixed models and Bayesian analyses from scratch, for always patiently answering my many questions, and for offering me feedback on my work at lightning speed. Hugh, thank you for encouraging me to have the confidence to push the boundaries of my research and teaching me not to be afraid of risky experiments.

Second, I am very grateful to the Economic and Social Research Council for generously offering me a studentship to fund my PhD (grant number ES/J500136/1), which allowed me to compensate participants, publish my research open access, and attend AMLaP 2016 and Joint Action Meeting 2017. I am also grateful to the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh for granting me several Research Support grants, which allowed me to attend AMLaP 2017 and compensate participants for Experiments 7-9.

Third, I would like to thank everyone who offered me technical assistance during this project. In particular, I thank Alice Turk for offering advice on analyzing the pitch contours of stimuli in Experiments 1-6, Sven Mattys for offering advice on rate manipulations for Experiments 5-6, Max Dunn for acting as second coder for audio files in Experiments 1-6, and Nige Corps for patiently (and with only minor protests!) recording the audio for all experiments in this thesis. I would also like to thank all the undergraduate students I have had the opportunity to work with. In particular, I am grateful to Abigail Crossley, who collected the data for Experiment 1 as part of her undergraduate dissertation.

Fourth, I am grateful to all my friends for providing me with endless opportunities to take my mind off work. Thank you to Wenjia for encouraging me to make bad decisions and always making my face hurt from laughing too much, to Laura for being my work wife and my banter soulmate, to Emma for pushing me to eat more cake, to Nick for the countless weeknight dinners, to Jordeen for chatting to me about everything and nothing, and to Rory for sending me memes to keep me going. Thank you also to my friends (Hayley and Katie) from back home for being there for spa days, shopping trips, and drives around town. Thanks also to the lunch crew (Guillermo, Michael, Milan). I probably would've been a lot more productive without you lot, but my life would've definitely been a lot less fun!

Finally, I cannot put into words (but I will try!) how grateful I am to all of my family for their endless support – you have always been interested in what I'm doing, even if it doesn't make much sense outside of academia. Mum, you have talked me through every PhD (and non-PhD) related issue imaginable and offered me great advice (even though I hate that you're always right!). Dad, you have always made me laugh with terrible (and sometimes impressive) jokes and have been very good at building most of the furniture in my flat! Josh you are always willing to put up with my weird sense of humor and (grudgingly) give me hugs. You three deserve a PhD just for putting up with me!

# Table of Contents

# Figures

# Tables

# 1. Literature Review[1]

During language comprehension, there is much evidence that listeners often predict what they are going to hear before they actually hear it. For instance, a listener who hears the utterance *Dogs are my favorite…* may predict that the speaker's likely next word is *animal*. These linguistic predictions are likely important during conversational dialogue, in which interlocutors take turns at talk with relatively little gap or overlap between their contributions (e.g., Clark, 1996). For example, if the listener can predict what the speaker is likely to say (a content prediction), then they may be able to use this prediction to prepare their own response. But listeners must not only predict *what* the speaker is going to say: They also need to predict *when* the speaker is going to finish (a timing prediction), so they can time articulation. Although it is well-documented that listeners predict content and timing during language comprehension, it is less clear what role these predictions play during conversational dialogue, in which people must deal with the additional demands of generating predictions in a timely manner. This thesis investigates how listeners use prediction to comprehend a speaker's turn and coordinate their utterances during conversational dialogue.

The current chapter first provides an overview of existing findings that support the notion of content prediction during language comprehension. This thesis is partly concerned with the prediction of lexical, syntactic, and word form

[1] Parts of this chapter are based on a manuscript published in *Discourse Processes* (Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes, 55,* 230-240.). Authorship details: Corps wrote the original manuscript and Gambi and Pickering acted as supervisors and contributed to the revision of the manuscript.

information, and so Section 1.1 focuses on studies investigating prediction of these three sources of information. Since this thesis is also concerned with the mechanisms responsible for predicting *when* events occur, we review evidence that suggests listeners make timing predictions during language comprehension (Section 1.2). After setting this background, we then consider the role of prediction in two important areas of language processing: conversational turn-taking (Section 1.3) and comprehending utterances under difficult conditions (Section 1.4). Specifically, Section 1.3 focuses on how listeners use content and timing predictions to prepare a response (Section 1.3.1) and time articulation (Section 1.3.2). Section 1.4 then considers how listeners may use prediction of word form to comprehend speech that is difficult to understand (e.g., when encountering a speaker with an unfamiliar accent).

## 1.1. Content Prediction during language comprehension

When comprehending an utterance, people typically process information incrementally (i.e., on a word-by-word basis). For example, Frazier and Rayner (1982) found that readers would rapidly adopt one interpretation when presented with ambiguous sentences (e.g., interpreting the phrase *on the cart* as attached to the verb *loaded* rather than to the noun phrase *the boxes* in the sentence *Sam loaded the boxes on the cart)*, suggesting that syntactic parsing is incremental. In addition, listeners tend to fixate objects immediately after hearing the relevant words in a sentence (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Although these studies demonstrate incrementality at the sentence level, it can also occur at the lexical level: Allopenna, Magnuson, and Tanenhaus (1998) showed that participants fixated pictures (both more often and for longer) whose name shared the initial or final phonemes with spoken target words relative to those with no phonological overlap.

However, listeners do not only process each word as they encounter it: They can also predict what the speaker is likely to say before they actually say it. For example, participants often converge on a continuation (e.g., *spoon*) when presented with sentence contexts such as *At the dinner party, I wondered why my mother wasn't eating her soup. Then I noticed that she didn't have a….* Importantly, this effect does not occur only in laboratory tasks. In natural conversations, interlocutors sometimes complete each other's utterances (e.g., Howes, Purver, Healey, Mills, & Gregoromichelaki, 2011), suggesting that the listener comprehends the speaker's incoming utterances, but also predicts what the speaker is likely to say next.

Prediction is thought to occur when the comprehender pre-actives linguistic information before they encounter the relevant input (e.g., Altmann & Kamide, 1999). As a result, the listener carries out some of the relevant processing in advance, which thus facilitates later comprehension. For example, if listeners predict *spoon* after hearing the sentences in the previous paragraph, then they will find it easier to process this word when the speaker actually produces it. Note that prediction contrasts with integration, which assumes that comprehension is facilitated simply because listeners find it easier to integrate predictable rather than unpredictable words into the preceding context (see Kutas, DeLong, & Smith, 2011). For example, the predictable word *spoon* is a more plausible fit to the context of the sentence in the previous paragraph than less predictable items, such as *fork*, which may make it easier to integrate. Thus, integration accounts can explain faciliatory effects when the input is actually processed, but they do not assume that the input is predicted.

But what information do people predict? In the sections that follow, we review evidence that suggests that listeners can predict a speaker's utterance at various linguistic levels (semantics, syntax, and form/phonology). Although there is a debate about the mechanisms that underlie content prediction (i.e., some theories argue that prediction is comprehension-based, while others argue that prediction is production-based; e.g., Pickering & Garrod, 2013; Dell & Chang, 2014), such a distinction is beyond the scope of this thesis and so we do not discuss these theories in detail.

### 1.1.1. Predicting semantics

Some research exploring prediction during language comprehension has used the *visual-world paradigm*, in which participants view a visual scene (usually consisting of many objects) while simultaneously listening to sentences. Predictive looking is thought to occur when listeners attend to an object before it is actually mentioned. In one of the first studies to use this method, Altmann and Kamide (1999; see also Kamide, Altmann, & Haywood, 2003) recorded participants' eye movements while they viewed visual scenes (e.g., a picture of a boy, a cake, a toy car, a toy train set, and a ball) and simultaneously listened to sentences. Sentences (e.g., *The boy will eat…*) could apply to only one object in the scene (e.g., *the cake*), thus making the mention of the cake predictable, or could apply to any of the objects (e.g., *The boy will move…*), making it impossible for the listener to confidently predict how the sentence would continue. When participants heard the verb *eat* they looked towards a picture of a cake earlier and for longer than when they heard the verb *move*, suggesting that they used the semantics of the verb to predict which of the objects was most likely to be mentioned next (i.e., edible objects).

In many studies, predictability is often assessed using an offline cloze task (Taylor, 1953; see also Staub, Grant, Astheimer, & Cohen, 2015), in which participants are presented with incomplete sentence fragments (e.g., *The boy will eat*…) and are asked to provide the word(s) that they think is most likely to follow. The cloze probability of each continuation is computed by determining the proportion of participants who provided a particular completion. When an utterance is predictable, cloze probability is high and participants tend to converge on a completion. When cloze probability is low, the utterance is considered unpredictable and participants' completions tend to differ.

The cloze task has been used to select stimuli for a number of electroencephalography (EEG) studies, in which participants are presented with predictable contexts followed by expected or unexpected continuations. For example, Federmeier and Kutas (1999) recorded event-related brain potentials (ERPs) while participants read discourse contexts that predicted a particular continuation (e.g., *They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of...*). These contexts were followed by (1) the predictable word (e.g., *palms;* average cloze probability of 75%), (2) a semantically related implausible word (e.g., *pines*), or (3) a semantically unrelated implausible word (e.g., *tulips*). The authors found N400 effects for the unexpected words, regardless of whether they were from the same or different semantic categories, compared to predictable words. But importantly, this N400 was reduced for the semantically related implausible words than for the semantically unrelated implausible words, suggesting that participants may have predicted the shared semantic category (e.g., TREE), leading to easier integration of palms than tulips.

In another study, Otten and Van Berkum (2008; Experiment 1a) presented participants with semantically inappropriate nouns (e.g., *stove*) embedded in predictive contexts (e.g., *Sylvie and Joanna really feel like dancing and flirting. Therefore they go to a* ***stove****, where they also make very nice cocktails)*, which predicted a particular noun (e.g., *disco*; average cloze of 65%), or non-predictive contexts (e.g., *After all the dancing Joanna and Sylvie really don't feel like flirting tonight. Therefore they go to a* ***stove*** *where they also have a nice and quiet chill-out zone*), which predicted any number of different words (e.g., *restaurant, hotel,* etc.; average cloze of 35%). They found that inappropriate nouns presented in a predictive context elicited a more positive ERP than those presented in a non-predictive context, suggesting that participants predicted the upcoming word (e.g., *disco*), and were surprised when a different word occurred instead. Together with eye-tracking studies, this research suggests that listeners can predict specific words in a speaker's utterance.

However, since these EEG effects occurred on or after the unexpected word, they could also reflect ease of integration: Less expected words are likely harder to integrate into the sentence than more expected words because they are less plausible continuations (e.g., Rayner, Warren, Juhasz, & Liversedge, 2004). Better evidence for prediction comes from Grisoni, McCormick, Miller, and Pulvermüller (2017), who presented participants with contexts that predicted face or hand-related continuations (e.g., *I take some grapes and I* ***eat*** or *I take the pen and I* ***write***). They found that participants activated body specific parts (e.g., face for *eat* and hand for *write*) of the motor cortex before the onset of the predicted continuation. This motor activation did not occur in unpredictable contexts, which could have been continued with any number of verbs (e.g., *I do not take the pen and I…*) Thus, these results suggest that participants predicted the semantics of the upcoming verb.

### 1.1.2. Predicting syntax

Some researchers have explored whether comprehenders can predict the syntactic structure of a speaker's utterance. For example, Staub and Clifton (2006) found that participants read the phrase *or the subway* faster after reading *the team took either the train…* than after *the team took the train….* The authors argued that this effect occurred because listeners predicted that a coordination structure would likely follow the word *either*, which facilitated processing of this structure when it actually occurred. When participants encountered *or the subway* after *the team took the train*, they had to reanalyze, thus leading to slower reading times.

In an EEG study, Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005; see also Otten, Nieuwland, & Van Berkum, 2007) presented participants with Dutch two-sentence discourses (e.g., *The burglar had no trouble locating the secret family safe. Of course it was situated behind a big but unobtrusive…*), which predicted a particular continuation (e.g., *painting*$_{\text{neu}}$*;* average cloze of 86%). These sentences continued with either the expected adjective or an unpredictable (but plausible) adjective that differed in syntactic gender (e.g., *bookcase*$_{\text{com}}$). Participants showed a larger differential ERP effect when the discourse contexts were continued with adjectives that mismatched the syntactic gender of the expected continuation. In a similar study by Wicha, Bates, Moreno, and Kutas (2003; see also Wicha, Moreno, & Kutas, 2003, 2004), native Spanish speakers listened to sentence contexts missing a critical word (e.g., *Red riding hood carried the food for her grandmother in a…. But the wolf arrived before she did)* and viewed a line drawing, which was either the expected continuation (e.g., *basket;* average cloze probability of 67%), or a semantically incongruent continuation of the same gender (e.g., *crown*). In half of these sentences, an article of the wrong gender preceded the drawing (e.g., *un* in Spanish; *un canasta* [basket]/*corona* [king]), which created a gender agreement violation. The authors found that articles with gender markings different from the gender of the expected noun elicited a larger negativity between 300 and 500 ms compared to articles of the expected gender. Together, these results suggest that listeners can pre-activate syntactic features of upcoming words.

### 1.1.3. Predicting word form

Other studies provide evidence to suggest that listeners can predict word form information. For example, DeLong, Urbach, and Kutas (2005) presented participants with high cloze sentences (e.g., *The day was breezy so the boy went outside to fly…*), which were followed by the predicted article completion (e.g., *a kite*) or an unpredictable, but equally plausible, completion (e.g., *an airplane*). Listeners displayed a larger N400 effect when indefinite articles mismatched rather than matched an unexpected upcoming noun, suggesting that participants predicted the form of the upcoming noun. In addition, the amplitude of this N400 effect varied as a result of the cloze probability of the predicted completion. In other words, when the sentence context was less strongly biased towards a specific completion (i.e., when cloze was low), the N400 amplitude in response to a mismatched indefinite article was also lower.

However, Nieuwland et al. (2017; see also Ito, Martin & Nieuwland, 2016) did not find these effects in a nine-lab replication study (but see DeLong, Urbach, & Kutas, 2017; Ito, Martin, & Nieuwland, 2017), suggesting that they may not be particularly consistent. Moreover, research suggests that more frequent phrases tend to be comprehended more quickly than less frequent phrases (e.g., Arnon & Snider, 2010), suggesting that common phrases are stored in the mental lexicon. Thus, it is possible that comprehenders may also store article-noun sequences, and so participants in DeLong et al. (2005) may have predicted these sequences instead of predicting word form. Finally, English articles are only informative about the initial phoneme of the next word. In other words, there is no phonological dependency between the article and the noun, and thus it is unclear why participants would use the form of the article to predict the upcoming noun (see Ito et al., 2017).

In another study, Laszlo and Federmeier (2009) recorded ERPs while participants read predictive sentences that were completed with an expected continuation (mean cloze probability of 89%), or an unexpected continuation (a word, pseudoword, or illegal string) that was either orthographically related or unrelated to the expected continuation. Participants showed a reduced N400 effect for unexpected items that were orthographically related to the expected continuation compared to items that were orthographically unrelated. Since all of the unexpected items had a similarly poor semantic fit to the context, Laszlo and Federmeier concluded that the N400 effect was associated with orthographic overlap and suggested that listeners predicted form information. However, it is possible that this pattern of activation occurred because orthographically related words are more easily integrated into the context than orthographically unrelated words. Predictive sentences may lead to the activation of the semantics of the expected word, which subsequently activates its orthographic form. This orthographic form then facilitates processing of the orthographic neighbours of the predicted word, which results in a reduced N400 for orthographically related items.

It is also possible that Laszlo and Federmeier's (2009) results reflected task-specific processing (e.g., Newman, Connolly, Service, & McIvor, 2003) especially given that participants were asked to judge whether each stimulus was a normal sentence. However, other studies that have used passive comprehension tasks have found evidence for form prediction. In a study by Ito et al. (2016), participants read high (e.g., *The juice isn't cold enough, so Alice is adding some…*) or medium (e.g., *The family went to the sea to catch some…*) cloze sentences, which were continued with the predictable word (*ice* in the high cloze example; *fish* in the medium cloze example), an anomalous word sharing form features with the predictable word (high cloze: *dice*; medium cloze: *wish*), an anomalous word that was semantically related to the predictable word (high cloze: *cube*; medium cloze: *pond*), or an unrelated anomalous word (high cloze: *wine;* medium cloze: *echo*). These sentences were presented via visual serial presentation at either a normal (300 ms word duration; 200 ms inter-word interval) or slow rate (500 ms word duration; 200 ms inter-word duration). At both presentation rates, anomalous words in all conditions elicited an N400 effect, but the N400 effect for semantically related words was reduced compared to unrelated words. When sentences were presented at a slow rate, the N400 effect for form related words was also reduced. Thus, these results participants can predict both the form and the meaning of a speaker's utterance when given enough time.

### 1.1.4. Conclusion

Although there may be controversy surrounding how listeners generate predictions during language comprehension, the literature reviewed in this section demonstrates that there is typically consensus that listeners can predict the semantics, syntax, and form of upcoming words. In Section 1.3, we discuss how semantic and syntactic predictions may help interlocutors coordinate their utterances during conversational turn-taking, and Experiments 1-4 (Chapter 2) explore these issues experimentally. Section 1.4 considers how listeners may use prediction of word form to comprehend distorted speech, and Experiments 7-9 (Chapter 4) investigate this issue experimentally.

## 1.2. Timing prediction during language comprehension

In the previous section, we reviewed evidence that suggests listeners can predict *what* a speaker is likely to say (i.e., the content of an utterance) during language comprehension. However, the timing of these utterances (e.g., the rate at which they occur) is also important for successful comprehension. For example, listeners use the speech rate of an utterance to identify phonemes (e.g., Port, 1979; Miller, 1981), perceive lexical stress (e.g., Reinisch, Jesse, & McQueen, 2011), and identify word boundaries (e.g., Turk & Shattuck-Hufnagel, 2000). Speech rate information is likely particularly relevant during conversational dialogue, in which interlocutors tend to vary considerably in their speaking rate (ranging from 3.45 to

5.45 syllables per second; Tauroza & Allison, 1990). This section reviews evidence that suggests listeners represent the speech rate of utterances and use this information to predict the rate of forthcoming speech.

Much evidence suggests that listeners entrain to (or track) an interlocutor's speech rate using cyclic neural oscillators, which are pools of neurons that synchronize to an external rhythm (Large & Jones, 1999). For example, Zion Golumbic et al. (2013) recorded electrocorticographic (ECoG) activity in the auditory cortex while listeners attended to one of two speakers. They found that oscillations in both the high (75-150 Hz; associated with phrasal processing; see Giraud & Poeppel, 2012) and low (1-7 Hz, associated with phonemic and syllabic processing) frequency ranges tracked the signal of the attended speech. In other words, there was a close correspondence between oscillatory activity and the speech signal. In follow-up analyses, higher frequency effects were shown to reflect evoked responses to the attended speech stream, while low frequency effects reflected processes related to speech perception (see also Ding & Simon, 2012; Mesgarani & Chang, 2012).

In a related study, Luo and Poeppel (2007; see also Ahissar et al., 2001) recorded magnetoencephalography (MEG) signals while participants listened to sentences that varied in their intelligibility. The authors found that low frequency oscillations (in the theta band; 4-8 Hz) tracked the speech signal. Additionally, tracking accuracy correlated with speech intelligibility, such that tracking was less accurate when sentences were less intelligible. Finally, Park, Ince, Schyns, Thut, and Gross (2015) found better coherence between signals in the speech stream and the auditory cortex for forward than backward speech.

Reduced oscillatory tracking of unintelligible speech may occur because oscillators are thought to be sensitive to the speaker's rate of syllable production (see Giraud & Poeppel, 2012). Fluctuations associated with syllabic rate are likely reduced when speech is unintelligible, thus leading to lower tracking accuracy (i.e., a lower correspondence between oscillatory activity and the speech signal). Indeed, some studies suggest that this may be the case. For example, Doelling, Arnal, Ghitza, and Poeppel (2014; see also Ghitza, 2012) found that the tracking accuracy of neural oscillators was reduced when fluctuations associated with syllable rate were removed (and intelligibility was reduced). Envelope tracking was regained when these fluctuations were artificially reinstated by inserting silent gaps, so that the syllable rate of the manipulated turn was comparable to that of the natural turn (and speech was intelligible).

Once listeners have entrained to their interlocutor's syllable rate, they can use this entrainment to predict the rate of the speaker's forthcoming utterance. For example, Dilley and Pitt (2010) either expanded (by a factor of 1.9) or compressed (by a factor of 0.6) the rate of the context surrounding a co-articulated single-syllable function word (e.g., *Deena doesn't have any leisure* ***or*** *time*). When context rate was expanded, listeners tended not to perceive a function word (e.g., *leisure or time* was perceived as *leisure time*); when context rate was compressed, listeners tended to erroneously perceive an absent function word (e.g., *leisure time* was perceived as *leisure or time*). These results are not only limited to function words, but can also occur with reduced syllables (Dilley, Morrill, & Banzina, 2013).

This effect is thought to occur because the listener entrains to the speaker's rate of syllable production and predicts that future syllables will continue to be produced at the same rate. Syllables incorrectly appear or disappear because the critical function word is processed at the (incorrect) predicted rate. In other words, when the context rate is slowed, the listener predicts that the next syllable (i.e., the function word *or*) will be produced at the same slow rate. When this function word is produced at a faster rate than predicted, it is still interpreted at the predicted slow rate, which leads to the loss of a syllable.

The results from research by Kösem et al. (2017) provide support for this interpretation. They conducted an EEG study, in which participants listened to Dutch sentences with varying speech rate. Specifically, the beginning of the sentences (the carrier window) was either presented at a fast or a slow rate, while the last three words (the target window) were presented at an intermediate rate. Participants were instructed to report the last word of the sentence, which contained an ambiguous vowel that could be interpreted as a short /*a*/ (e.g., *tak* or "branch") or a long */a*:/ (e.g., *taak*, or "task"). Much like Dilley and Pitt (2010), Kösem et al. found that the speech rate of the carrier window influenced the perception of the target word: The behavioral results indicated that participants tended to perceive a word with a long vowel (e.g., *taak*) after a fast speech rate, and a word with a short vowel (e.g., *tak*) after a slow speech rate. In addition, magnetoencephalography (MEG) analysis of the auditory cortices showed that low frequency activity entrained to the speech rate of the carrier window. This entrainment was sustained in the target window and correlated with behavioural performance. In other words, participants entrained to the rate of the carrier window, which led them predict that the target window would be produced at the same rate.

Other research suggests that timing predictions based on speech rate are not limited to the immediately preceding sentence frame, but can also build up over the course of multiple utterances. In one study, Baese-Berk, Heffner, Dilley, Pitt, and McAuley (2014) manipulated the speech rate of individual utterance frames (the distal rate) and the average speech rate of utterances across the whole experiment (the global rate). They replicated Dilley and Pitt's (2010) earlier results, and found that participants were less likely to perceive a function word when the context rate of an individual utterance was slowed. In addition, listeners were less likely to perceive a function word when the global speech rate was slower. Together, these results suggest that listeners can make timing predictions across multiple timescales (i.e., both over the course of an individual utterance and over many utterances).

In sum, there is much evidence to suggest that listeners can entrain to a speaker's syllable rate and can subsequently use this entrainment to predict the rate of the speaker's forthcoming syllables. Section 1.3.2.3 reviews evidence that suggests entrainment plays a role in conversational dialogue and considers how listeners could use syllabic entrainment to time articulation of their own turns during conversational dialogue.

## 1.3. Conversational turn-taking

In the previous sections, we reviewed evidence that suggests listeners can predict content and timing during language comprehension. Such predictions can ease cognitive processing and help listeners get ahead of the game. But in addition, content and timing predictions may be particularly useful during conversational dialogue, which is arguably the most basic form of language use.

During conversation, interlocutors repeatedly and regularly switch between comprehending their partner's utterance and producing an appropriate and timely response. These processes are so finely coordinated that there is often little gap or overlap between turns. Indeed, Stivers et al. (2009) found average inter-turn intervals between 0 and 200 ms in a comparison of ten different languages, with overlap occurring only about 5% of the time (Levinson, 2016). In contrast, research suggests that language production is much slower, with a single word taking between 600 and 1200 ms to produce, depending on word frequency (Indefrey & Levelt, 2004; Levelt, Roelofs, & Meyer, 1999), and a complete utterance taking around 1500 ms (Ferreira, 1991; Griffin & Bock, 2000).

Current theories agree that interlocutors achieve such timings using prediction (e.g., Garrod & Pickering, 2015; Bögels & Levinson, 2017). Some research has focused on how listeners can use such predictions to articulate their response at the appropriate moment, so they do not overlap with the previous speaker (e.g., Magyari, Bastiaansen, De Ruiter, & Levinson, 2014). But if listeners can predict what the speaker will say before the speaker reaches the end of their turn, then the listener may also be able to begin preparing their own response in advance of the turn-end (e.g., Bögels, Magyari, & Levinson, 2015), which will ease some of the timing burden from the language production system. In the sections that follow, we discuss how prediction may help listeners time articulation and prepare a response during conversation and consider how these processes may be interwoven.

### 1.3.1. Timing response articulation

To ensure smooth conversational dialogue, listeners must appropriately time articulation of their own turn so they do not extensively overlap with the current speaker. Although much research has explored the mechanisms that listeners use to time articulation, it is not currently clear how they do so.

One possibility is that listeners react to the presence of linguistic (e.g., drawl on the final syllable of the utterance) and non-linguistic (e.g., termination of hand gestures) turn-final cues, which signal that the utterance is coming to an end. According to this *reactive* account (e.g., Duncan, 1972), listeners do not use prediction to time response articulation. This contrasts with a *turn-end prediction* account (e.g., Sacks, Schegloff, & Jefferson, 1974), which assumes that listeners time articulation by predicting (or projecting) when the speaker will reach the end of their utterance. Listeners are thought to determine this moment by predicting the lexical and/or the syntactic content (i.e., what the speaker is going to say) of the turn.

Although the majority of research has focused on contrasting the reactive and turn-end prediction accounts, a third possibility is that listeners predict when they should launch response articulation using timing predictions based on speech rate entrainment (e.g., Garrod & Pickering, 2015; Wilson & Wilson, 2005). But even though research has demonstrated predictive timing during comprehension (see Section 1.2), very little has investigated whether timing predictions influence the timing of articulation.

The following sections review literature that suggests listeners may use each of these mechanisms to time articulation of their turns during dialogue. Since the primary focus of this thesis is on predictive mechanisms, Section 1.3.1.1 only briefly reviews the literature on turn-final cues, to make it clear why we controlled for their presence in Experiments 1-6. Section 1.3.1.2 reviews evidence for turn-end prediction, and discusses which information may help listeners determine when the speaker will reach the end of their utterance. This section provides some of the background theory and literature for Experiments 1-4 (Chapter 2). Finally, Section 1.3.1.3 reviews evidence that suggests listeners can time articulation using speech rate entrainment and sets the theoretical background for Experiments 5 and 6 (Chapter 3).

#### 1.3.1.1. Turn-final cues

Duncan (1972, 1974; Duncan & Niederhe, 1974) proposed that listeners initiate response articulation after the speaker displays turn-final cues, which signal that they wish to yield their turn. Using transcriptions of two dyadic interviews, Duncan identified six possible linguistic and non-linguistic cues that may be used to time response articulation: (1) drawl on the final syllable of the utterance; (2) a drop in pitch and/or intensity; (3) falling or rising phrase final pitch; (4) the completion of a grammatical clause; (5) the termination of hand gestures[2]; and (6) using sociocentric sequences, such as "but uh" or "you know", which do not add any substantive information to the speech context. Interlocutors were less likely to produce overlapping talk when they attempted to take a turn after the speaker

[2] Note that other visual turn-final cues have also been proposed (e.g., gaze direction; Kendon, 1967). But since all experiments in this thesis involved only auditory stimuli, we do not discuss these further. Interestingly, there are some reports that inter-turn intervals may be similar in telephone and face-to-face interactions (e.g., De Ruiter et al., 2006). Thus, visual cues may not be necessary for turn-end prediction (see also Gambi, Jachman, & Staudte, 2015).

displayed a turn-final cue (see also Local, Kelly, & Wells, 1986; Local & Walker, 2012). In addition, listeners were more likely to make a turn-taking attempt when the speaker displayed more turn-final cues.

However, there are a number of notable issues with Duncan's findings (for more detailed criticisms, see Beattie, 1981; Cutler & Pearson, 1986). First, the results are correlational, and so we cannot infer the direction of causality: The observation that certain turn-final cues co-occurred with speaker switches is not evidence that listeners actually used these cues to time articulation. Relatedly, this correlational analysis was based on two dyadic interviews, in which speakers displayed five-turn final cues simultaneously in only nine instances (note that there were no instances in which speakers displayed the maximum of six turn-final cues), and so it is unclear whether these results are representative of natural conversation.

Despite these issues, further experimental studies have demonstrated that listeners are indeed sensitive to the presence of turn-final cues. In one study, Cutler and Pearson (1986) created dialogue fragments by asking speakers to read written scripts, which contained utterances that occurred either at the end of a conversational turn (and should thus contain turn-final cues) or in the middle of a turn (and should contain turn-medial cues). The authors found that utterances judged as turn-final by a separate group of participants were associated with pitch downstep, which occurs when the next syllable of an utterance is significantly lower in speech than the previous syllable. In contrast, utterances judged as turn-medial were associated with a pitch upstep, which occurs when the next syllable is higher in pitch than the previous syllable. However, many of the utterances that listeners found ambiguous (i.e., those on which they could not agree on turn-final or turn-medial judgements) were also characterized by pitch downsteps or upsteps, suggesting that other cues must also play a role in determining whether an utterance is turn-final. In addition, listeners in this study did not have to produce a verbal response, and so the results do not necessarily demonstrate that these cues play a role in timing articulation.

In another study, Beattie, Cutler, and Pearson (1982) presented participants with extracts of turn-final (turns with a successful speaker switch), turn-medial (turns with no speaker switch), and turn-disputed (turns immediately preceding an interruption) utterances (consisting of at least one sentence) from television interviews with Margaret Thatcher. They identified five turn-yielding cues (e.g., a pitch downstep, a fall in pitch, whispery voice, creaky voice, and a quickness in tempo), which were present in the turn-final utterances more often than in the turn-medial utterances. The turn-disputed stimuli, in contrast, contained conflicting cues (e.g., they were characterised by a fall in pitch, but a fall that did not descend as low as turn-final utterances), which may have led to interruption in the original interviews. Together with Cutler and Pearson (1986), these results suggest that listeners are sensitive to turn-final cues, which they can use to determine whether the speaker wishes to relinquish their turn and to subsequently time response articulation. As a result, we made sure to control for the presence of turn-final cues in Experiments 1-6 to ensure that our results could not be attributed to differences in the occurrence of these cues.

However, these cues do not account for all successful speaker switches. In a corpus study of twelve dyadic task-oriented interactions, Gravano and Hirschberg (2011) assessed the role of seven turn-final cues identified by Duncan (1972) and found that they were significantly more likely to occur in stretches of speech preceding speaker changes than in those preceding a continuation of the current speaker's turn. But listeners were only 65% likely to take a turn when all seven cues were present, leaving open the possibility that other mechanisms (i.e., turn-end prediction) are also at play. Thus, the following section reviews the literature that suggests listeners can use turn-end prediction to time response articulation.

#### 1.3.1.2. Turn-end prediction

In one of the first accounts to suggest that listeners can predict turn-endings, Sacks et al. (1974) argued that listeners predict the lexico-syntactic content of the speaker's turn (i.e., they predict which turn constructional unit the speaker is using; e.g., whether a turn is a word, phrase, or clause) and then use this prediction to judge when the turn is likely to end. For example, if Laura had just asked Rory *"What would you like for your birthday*?" and Rory's reply began "*The…*", then Laura might predict that the question requires Rory to identify an object. The syntactic unit best suited to this purpose is a noun phrase, and so Laura will assume that Rory's turn will end as soon as he completes his noun phrase (see Power & Martello, 1986, for a similar example).

But since just about any phrase could constitute a turn constructional unit (e.g., a single word could be a turn on its own, such as *What?*, or could be part of a larger unit, such as *What is your favourite animal*?), Sacks et al. (1974) argue that listeners also use intonation to help them predict turn-endings. Note, however, that the role of intonation in the turn-end prediction account is different from that proposed by the reactive account. The reactive account assumes that prosodic turn-final cues signal immediate turn-ending, meaning that the listener could not have predicted the turn-end before it actually occurred (see Section 1.3.1.1). The turn-end prediction account, in contrast, assumes that intonation can be used to predict when the turn will end, such as whether the turn constructional unit will be continued by one or many words (see Grosjean, 1983). In other words, intonation is used to predict the length of the turn, rather than to detect its immediate end.

Experimental work has investigated this issue in more detail. In one study, De Ruiter, Mitterer, and Enfield (2006) assessed turn-end prediction using a button-press paradigm, in which participants listened to full turns taken from natural conversation and pressed a button when they expected the speaker to reach the end of their utterance. The authors either removed the words from the utterance using low-pass filtering (which leaves prosody unaltered) or set the pitch to a constant level (which leaves lexico-syntactic information unaltered). When pitch was flattened, participants responded on average 200 ms before the end of the speaker's turn, which was similar to the timing of verbal responses in the original conversations and the button-press responses to unmodified turns extracted from those conversations. When lexical information was removed, however, participants responded on average 500 ms before the end of the utterance. Although it is possible that other sources of prosodic information are important (e.g., final syllable duration; see Bögels & Torreira, 2015), these results nevertheless suggest that the actual words of the speaker's utterance are necessary for predicting turn ends. Indeed, additional research suggests that lexico-syntactic information is generally more important for turn-end prediction than intonation (see Lammertink, Casillas, Benders, Post, & Fikkert, 2015; Keitel, Prinz, Friederici, von Hofsten, & Daum, 2013).

But how do the speaker's words help listeners predict when the speaker will reach the end of their turn? In one study, Wesselmeier, Jansen, and Müller (2014) presented participants with turns containing semantic (e.g., *The priest always* ***grinned*** *the bell three times before he went to dinner*) or syntactic violations (e.g., *The priest always* ***rings*** *the bell three times before he went to dinner*). Using EEG, they measured the time course of Readiness Potentials (RP), which are associated with movement preparation, while participants completed the button-press task. Although there was no difference in button-press times between turns that contained semantic or syntactic violations and those that did not (the control utterances; e.g., *The priest always* ***rang*** *the bell three times before he went to dinner*), RPs were disrupted in the semantic and syntactic violation turns compared to control utterances. For the control sentences, participants displayed a RP around 1400 ms before the button-press; for the sentences with semantic or syntactic violations, the RP started around 900 ms before the button-press. They argued that their results suggest listeners use both semantic and syntactic information (provided by the speaker's words) to predict the turn-end. However, the syntactic error *rings* violates the tense of the sentence (i.e., it should be *rang*), and so participants must process the semantics of this word (at least to some extent) to detect the violation. Thus, it is possible that there was no difference in RPs to semantic and syntactic violations because they both required equivalent semantic processing to detect the error.

Nevertheless, additional studies suggest that semantic information may be more important for turn-end prediction than syntactic information. Using the button-press paradigm, Riest, Jorschick, and De Ruiter (2015; Experiment 3) found that listeners could still predict the speaker's turn-end when closed class words (which primarily serve a syntactic role; e.g., Brown, Hagoort, & Ter Keurs, 1999) were removed using low pass filtering, but not when open class words (which primarily serve a semantic role) were removed. But participants were most accurate at predicting the turn-end when both sources of information were available, suggesting that even though semantic information may be more important than syntactic information, both sources of information are necessary for turn-end prediction.

Together, the studies reviewed thus far demonstrate that listeners use lexico-syntactic content to predict the speaker's turn-end. However, these studies do not demonstrate that turn-end prediction is better when the semantic content or syntactic structure of the speaker's turn is more predictable. In other words, they do not demonstrate that listeners predict lexico-syntactic content and then use this prediction to determine the turn-end. Additional research has confirmed the importance of content predictability for turn-end prediction. In one study, Magyari et al. (2014) manipulated the content predictability of their stimuli, so that participants either could or could not predict what the speaker would say. In a gating paradigm, participants were auditorily presented with turns from actual conversations in fragments of increasing duration and were instructed to complete these turns with the words they expected to follow given the preceding context (much like a typical cloze task; Taylor, 1953). The authors assessed the predictability of these responses using entropy, which measures the consistency of completions across participants. Participants provided more consistent completions in the predictable (e.g., *I live in the same house with four women and another man*) than unpredictable condition (e.g., *She was again alone in the north*)[3], and were also more likely to complete predictable than unpredictable fragments with the words the original speaker had used. A separate group of participants, who completed the button-press task, responded before the end of predictable turns but after the end of unpredictable turns. Furthermore, concurrent EEG recordings showed a power decrease in the beta band at least 1250 ms before the end of the predictable but not the unpredictable turns.

In another study, Riest et al. (2015; Experiment 1) explored the role of content predictability by scrambling the word order of turns, so that participants could not use the preceding words of the speaker's turn to predict subsequent words. They found that participants responded around 300 ms before the turn-end when word order was scrambled, compared to 150 ms before the turn-end when participants heard the natural turn. Together with Magyari et al. (2014), these results suggest that listeners predicted the speaker's turn-end by predicting the content of the speaker's forthcoming utterance.

However, these studies have typically conflated measures of lexico-semantic content and syntactic predictability. Previous research suggests that listeners can predict the syntactic structure of the speaker's turn (e.g., Staub & Clifton, 2006; see Section 1.1.2). Thus, listeners may also be able to predict the speaker's turn-end even when they cannot predict the specific words the speaker will use (i.e., even when they cannot predict semantic content). Indeed, utterances can often be predictable in length but unpredictable in lexico-semantic content. To illustrate, the sentence fragment *Most people have two…* can be completed with many single words (e.g.,

[3] Note that we do not know where these fragments were cut off in the gating paradigm, or which of these words were provided as completions, since Magyari et al. (2014) do not provide this information.

*cars, dogs, siblings*), which overlap very little in their semantic content. Conversely, utterances can be unpredictable in length but predictable in content. For example, the sentence fragment *The Titanic sank after…* can be completed with *it hit an iceberg, hitting an iceberg,* or *crashing*, which all differ in length but overlap in content.

Only one study has investigated whether listeners can predict the word length of speakers' utterances. Using the same gating paradigm as Magyari et al. (2014), Magyari and De Ruiter (2012) assessed the number of words participants expected to complete sentence fragments. They found that the accuracy of turn-end prediction in De Ruiter et al.'s (2006) study correlated not only with the turn's content predictability (as in Magyari et al., 2014), but also with its length predictability (in number of words). More specifically, turns that elicited later button-presses tended to be completed with more words in the gating paradigm, while turns that elicited earlier button-presses tended to be completed with fewer words. Although such correlational data should be interpreted with some caution, these results suggest that listeners may also predict the turn-end by predicting the number of words the speaker will use. However, this study does not tell us whether predictions of semantic content can be dissociated from predictions of syntactic structure. In other words, can listeners predict response timing independently from predicting the semantic content of the speaker's turn?

In sum, studies exploring turn-end prediction suggest that both semantic predictability (i.e., predictions of what the speaker is going to say) and syntactic predictability (i.e., predictions of how many words the speaker will use) may play a role in turn-end prediction. However, these studies have not clearly established whether predictions of turn length can be made independently from semantic content.

In other words, there may be instances in conversation where listeners can use predictions of syntactic structure (i.e., turn length) to predict the turn-end, even when they cannot predict the semantic content of the speaker's turn. Conversely, there may also be instances where listeners can predict semantic content but cannot predict turn length. Thus, exploring this issue is relevant for understanding the information that listeners use to predict turn-endings and to time response articulation. Experiments 1-4 in this thesis (Chapter 2) address this issue by investigating whether listeners can predict the speaker's turn-end using predictions of turn length independently from predictions of turn content.

#### 1.3.1.3. Speech rate entrainment

A number of studies suggest that speech rate entrainment during comprehension can influence the rate of subsequent speech production. For example, Jungers and Hupp (2009; see also Jungers, Palmer, & Speer, 2002; Ten Bosch, Oostdijk, & Boves, 2005) presented participants with priming sentences produced at a fast or a slow rate. The authors found that when participants later produced picture descriptions, they were more likely to produce a response at a fast rate after hearing a prime at a fast rather than a slow rate, suggesting that their rate of production was influenced by the rate of the prime sentence. Similar results have been demonstrated in dialogue. Schultz, O'Brien, Phillips, and McFarland (2016) found that interlocutors' beat rates became mutually entrained during scripted turn-taking conversations: Participants produced their turn at a faster beat rate after their interlocutor produced their own turn at the same beat rate. In another study, Street (1984) found that interlocutors converged on both the speech rate and the duration of their turn transitions during dialogue. Together, these findings suggest that listeners entrain to their interlocutor's speech rate, which can in turn influence the rate of the listener's subsequent production.

However, these studies have not investigated whether speech rate entrainment influences the timing with which listeners initiate articulation during dialogue (i.e., the duration of the inter-turn interval). Some recent theories suggest that listeners not only use entrainment to predict the rate of the speaker's forthcoming syllables as they listen (see Section 1.2), but also to time response articulation according to the syllable rate of the speaker's turn. For example, Wilson and Wilson (2005) argued that each interlocutor's readiness to initiate syllable production rises and falls in cycles over the course of the conversation. At the peak of this oscillatory cycle, the speaker is maximally ready to produce a syllable. This readiness decreases until the mid-point of the speaker's syllable, after which readiness again begins to rise. Interlocutors' oscillatory cycles are in anti-phase, so that the listener's (as the next speaker) readiness to initiate a syllable is at a maximum when the speaker's is at a minimum (and vice versa), which may explain why conversational overlap is rare. In the context of turn-taking, anti-phase means that listeners will be maximally ready to produce their turn half a syllable before or after the end of the current speaker's turn. If the listener does not produce a response at this moment, then they will not be able to begin speaking again until after they have completed another oscillatory cycle (i.e., the duration of another syllable).

Although support for Wilson and Wilson's (2005) account can be drawn from studies demonstrating convergence of speech rate (e.g., Jungers & Hupp, 2009) and inter-turn intervals (e.g., Street, 1984), others have found that speech rate convergence does not influence inter-turn intervals (see Finlayson, Lickley, & Corley, 2012). Furthermore, there is very little evidence to support Wilson and Wilson's argument that interlocutors' oscillatory cycles are in anti-phase. In one study, Beňuš (2009) tested the oscillator theory using data from the Columbia Games Corpus of 12 dyadic conversations between speakers playing joint computer games. If interlocutors' oscillatory cycles are in anti-phase, then the listener should be equally likely to begin speaking half a cycle before or after the end of the speaker's turn, and so turn intervals should be bimodally distributed around zero. However, Beňuš did not find results consistent with this prediction. Instead, turn intervals were unimodally distributed, with a peak around 100-200 ms.

In another oscillator-based account, Garrod and Pickering (2015) also argued that the speaker's rate of syllable production influences the timing of the listener's subsequent syllables. Much like Wilson and Wilson (2005), Garrod and Pickering's account proposes that speech rate entrainment affects the duration of inter-turn intervals. Specifically, the authors argue that listeners use syllabic entrainment to predict the rate of the speaker's forthcoming syllables and the moment when they can launch articulation. As a result, turn transitions should be shorter when the speaker's syllable rate is faster rather than slower, because listeners should predict that they can launch articulation earlier. Research demonstrating that listeners can use speech rate entrainment to predict the rate of upcoming syllables (e.g., Dilley & Pitt, 2010) is consistent with this account.

However, research on predictive entrainment has focused solely on comprehension (see Section 1.2), and so it is unclear whether timing predictions based on speech rate entrainment during comprehension can influence the timing of response articulation. In other words, we do not know whether timing representations are shared across comprehension and production. Experiments 5 and 6 (Chapter 3) investigate this issue using a manipulation similar to Dilley and Pitt (2010) to test whether predictions based on speech rate entrainment influence the timing of response articulation.

#### 1.3.1.4. Conclusion

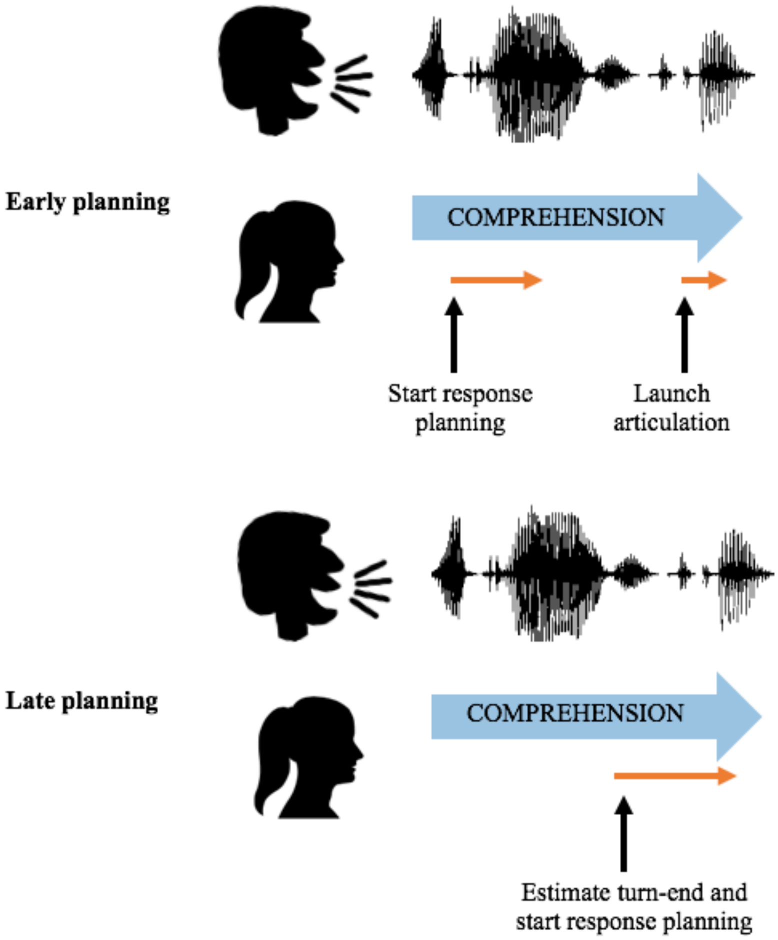
In sum, previous research suggests that listeners can use a number of different mechanisms to time response articulation during conversational turn-taking. Since this thesis is concerned with predictive timing, the subsequent studies focus on turn-end prediction and speech rate entrainment. Specifically, Experiments 1-4 (Chapter 2) investigate how listeners use semantic and syntactic predictions to determine the speaker's turn-end. Experiment 5 and 6 (Chapter 3) extend research in language comprehension on predictive entrainment and investigate whether timing predictions based on syllabic entrainment influences the timing of response articulation.

### 1.3.2. Response preparation

After having heard or predicted a sufficient part of the speaker's utterance, listeners can begin preparing their own response. Most theories of language production agree that preparation involves at least three stages: Message construction (conceptualization), formulation (lexical selection, structure building, and phonological encoding) and articulation (Bock, 1995; Levelt, 1983). But when do listeners begin preparing their response? Answering this question is important for understanding how listeners use predictions of content and timing during conversational turn-taking.

One possibility is that listeners prepare their response early in the speaker's turn and then hold this response in a buffer until they are given the opportunity to launch articulation. In other words, this *early-planning hypothesis* assumes that listeners use content predictions (i.e., predictions of what the speaker is going to say) to prepare the content of their own response independently from launching articulation (e.g., Levinson & Torreira, 2015; see Fig. 1). Listeners may then time response articulation either by predicting the speaker's turn-end, by reacting to turn-final cues, or a combination of the two (see Section 1.3.1). Early preparation may be advantageous because it relaxes some of the timing constraints of producing turns in a timely manner. However, language production is cognitively demanding (e.g., Roelofs & Piai, 2011) and so preparing and buffering a response could interfere with simultaneous comprehension. Importantly, listeners could minimize such interference by beginning response preparation only when they are sure they will soon have the opportunity to launch articulation. This *late-planning hypothesis* assumes that listeners do not prepare the content of their response as soon as they can predict what the speaker is going to say. Instead, preparation depends on predicting when they can time articulation of their response (i.e., content preparation depends on timing).

Figure 1. Models of response planning An illustration of the early and late planning models adapted from Bögels and Levinson (2017). Blue arrows represent comprehension processes. Orange arrows represent production processes.



The following sections focus on existing evidence for and against both of these hypotheses, and set some of the theoretical background for Experiments 1-4 (Chapter 2). Note that we limit the majority of our discussion to research on the timing of the start of response preparation (i.e., whether listeners prepare a response as soon as they can predict turn content, or whether such preparation depends on predicting response timing) and do not extensively consider what aspects or how much of their response the listener actually prepares, since this is beyond the scope of this thesis.

### 1.3.2.1. Evidence for early planning

Research exploring the time course of response preparation has used a variety of different methods. In one study, Bögels et al. (2015) measured EEG correlates during a question-answering task, in which the information (here *007*) needed for response preparation was available either early (e.g., *Which character, also called 007, appears in the famous movies*?) or late (e.g., *Which character from the famous movies is also called 007?*) in the utterance. Participants were quicker to answer when the critical information was available early rather than late, and EEG correlates revealed (i) a positive ERP effect in the middle frontal and precentral gyri, which overlap with brain areas involved in speech production (Indefrey & Levelt, 2004), and (ii) reduced alpha power, which is associated with motor response preparation (Babiloni et al., 1999). Both of these effects occurred around 500 ms after the onset of the critical information necessary for response preparation, suggesting that listeners prepared their own response as soon as they could predict the content of their answer. Thus, these results suggest that the processes of content prediction and response preparation can be decoupled from timing articulation. After hearing *007*, listeners can predict the speaker's intention (e.g., that the question is likely to be related to James Bond) and can prepare a response consistent with this prediction, even though they do not know when they will have the opportunity to articulate this response.

However, we note that Bögels et al. (2015; see also Bögels, Casillas, & Levinson, 2018) used general knowledge questions, and so answers likely had to be retrieved from episodic memory. Although previous experimental research has found that the middle frontal and precentral gyri are associated with language production processes (Indefrey & Levelt, 2004), other studies report that the middle frontal gyrus may also be involved in episodic memory retrieval (e.g., Cabeza, 2002; Rajah, Languay, & Grady, 2011; Raz et al., 2005). Even though Bögels et al. did not find the same pattern of activation in a control study, in which participants memorized the questions, their results may still reflect the processes of retrieving the necessary answer from memory. Of course memory retrieval is necessary for conceptualization (i.e., participants would not be able to prepare their response without retrieving the relevant memory trace), but it is not clear whether Bögels et al.'s findings *only* reflect memory retrieval processes associated with conceptualization or whether they *also* reflect later stages of preparation.

Nevertheless, additional research using other tasks has found converging evidence for early response preparation. Barthel, Sauppe, Levinson, and Meyer (2016; see also Barthel, Meyer, & Levinson, 2017) used a task in which German participants completed a confederate's pre-recorded utterances. Since participants had to name any on-screen objects that the confederate had not already named, participants could (in principle) plan their response as soon as the confederate began uttering their last object name (indicated by the use of the word *and*; e.g., *I have a door and a bicycle*). The authors also manipulated the predictability of the confederate's turn-end, so that participants could or could not predict that a sentence final verb would follow the last object name. Both eye-movements and response latencies suggested that participants planned their response as soon as possible. However, neither of these measures were influenced by the predictability of the speaker's turn-end, suggesting that preparation did not depend on an accurate turn-end prediction. Thus, they conclude that participants prepared their response early, independently from launching articulation. However, it is possible that any turn-end predictions may have been overridden by the processes of response preparation, especially since participants could not launch articulation (i.e., indicate the turn-end) without having prepared their response (see Section 1.3.1.2 for a review of more explicit tasks assessing turn-end prediction independently of response preparation).

In instances where listeners prepare their response early, they must need to store this response in a buffer until it can be articulated. Results from immediate and delayed picture-naming studies, in which participants name pictures while ignoring distractor words, suggest that participants can buffer their utterances at various stages of production (e.g., Mädebach, Oppermann, Hantsch, Curda, & Jescheniak, 2011; Piai, Roelofs, & Schriefers, 2011; Piai, Roelofs, & Schriefers, 2014; Schriefers, Meyer, & Levelt, 1990). For instance, Piai et al. (2011) found that participants were slower to name pictures when distractor words were semantically related (known as the *semantic interference effect*) in an immediate but not in a delayed naming condition. In the immediate condition, a semantically related distractor word interfered with ongoing lexicalisation. No interference occurred in the delayed condition, however, because participants had most likely already completed the processes of lexical selection. In these instances, it is possible they were buffering their response at the phonological level until they could launch articulation.

Consistent with this argument, Piai, Roelofs, Rommers, Dahlslätt, and Maris (2015a) found alpha-beta desynchronization (8-30 Hz) in the occipital cortex and beta synchronization (12-40 Hz) in the middle frontal and superior frontal gyri during delayed but not immediate naming. Alpha-beta desynchronization has been associated with motor aspects of articulation (Piai, Roelofs, Rommers, & Maris, 2015b), while beta synchronization has been associated with maintaining the current cognitive state until the response can be articulated (Engel & Fries, 2010; Kilavik, Zaepffel, Brovelli, MacKay, & Riehle, 2013). These findings suggest that if listeners prepare their response in advance of articulation, they buffer and continue to rehearse this response, presumably so they do not forget what they wish to say, until they are given the opportunity to take their turn.

In sum, the studies reviewed in this section have explored whether listeners can use predictions of turn content to prepare their own response early in the speaker's turn, before they will have the opportunity to launch articulation. When they do prepare their response in advance, listeners can hold this response in an articulatory buffer until they are given the opportunity to launch articulation. However, preparing a response and holding it in an articulatory buffer may interfere with the listener's ability to concurrently comprehend their interlocutor's incoming turn, which may in turn interfere with their ability to predict response timing. The next section discusses these issues in more detail.

#### 1.3.2.2. Problems with early planning

In instances where the listener prepares their response early, they must represent both their prepared response (using production mechanisms) and their interlocutor’s utterance (using comprehension mechanisms). Previous neural studies suggest that production and comprehension recruit overlapping neural circuits (e.g., Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert, Menenti, Weber, Petersson, & Hagoort, 2012; Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Watkins, Strafella, & Paus, 2003; Wilson, Saygin, Sereno, & Iacoboni, 2004) and thus most likely share resources. For example, Segaert et al. found that the same brain areas (the left inferior frontal gyrus, the left middle temporal gyrus, and the bilateral supplementary motor area) were sensitive to syntactic repetition during comprehension and production. As a result, using production mechanisms to prepare and buffer an early response may interfere with the concurrent process of comprehending the speaker’s turn.

Indeed, numerous picture-word interference (PWI) experiments, in which participants name pictures while listening to or reading distractor words, have shown that participants are slower to name pictures in the presence of words (even when the words are unrelated) than pseuwodords (e.g., Dhooge & Hartsuiker, 2012), noise (e.g., Schriefers et al., 1990), or strings of X’s (Glaser & Glaser, 1982, 1989). In other words, comprehension interferes with simultaneous speech planning. However, it is unclear whether the inverse relationship holds, that is whether response preparation interferes with comprehension.

In one study investigating this issue, Jongman and Meyer (2017) used a picture-naming task, in which half of the participants named the picture while the other half listened to a pre-recorded speaker name the picture (i.e., planning condition was manipulated between-participants). In addition, pictures were preceded by auditory primes which were either identical to, associatively related to, or unrelated to the target picture. The authors found fastest naming latencies for pictures preceded by an identity prime, intermediate latencies for those preceded by an associatively related prime, and slowest latencies for those preceded by an unrelated prime. This priming pattern was the same regardless of whether participants named the non-target picture, suggesting that speech planning did not interfere with concurrent comprehension of the prime. Jongman and Meyer replicated the identity priming effect in a second experiment, in which participants had to decide whether or not to name the picture at the start of each trial (i.e., planning condition was manipulated within items). However, in this experiment they found an associative priming effect only when participants did not have to name the picture, suggesting that response preparation interfered with comprehension. The lack of effect of associative priming in the planning condition was likely related to the difficulty of the task. In Experiment 1, participants' task was predictable and they knew whether they would need to plan a response before picture onset. In Experiment 2, however, participants had to switch between planning and listening, which was likely cognitively demanding. This is particularly relevant for natural conversation, since the cognitive load is likely to be greater than in Jongman and Meyer's task, given that participants often have to prepare (and comprehend) a longer, more complex (e.g., multi-word response).

In another study, Bögels et al. (2018) instructed participants to complete the same question-answering task used by Bögels et al. (2015), but they also simultaneously viewed two pictures on-screen (e.g., a banana and a pineapple). Much like the previous study, the information (here *curved*) necessary for response preparation was available either early (e.g., *Which object is curved and is considered* *to be a type of fruit*?) or late (e.g., *Which object is considered to be a type of fruit and is curved*?). But in addition, the questions contained either an expected or unexpected word (e.g., *healthy* rather than *fruit* in both examples). The authors found that participants responded later to questions with an unexpected rather than expected word regardless of when critical information became available, suggesting that listeners still comprehended these words even when they planned their response early. In addition, an N400 effect occurred at the unexpected word in both the early and late planning conditions. However, the size of this N400 effect varied as a result of participants' response latencies: Participants with slower response times showed a larger N400 effect than those with faster response times. Together, these results suggest that fast responders allocated less resources to comprehension (leading to a smaller N400 effect) when they encountered the information necessary for response preparation. In contrast, slow responders allocated more resources to comprehension (leading to a larger N400 effect). Thus, this study provides some preliminary evidence that response preparation can interfere with concurrent comprehension.

Importantly if the degree of interference between preparation and comprehension is sufficiently large to be problematic, then listeners may instead prepare a response late in the speaker's turn, when they are sure they will soon have the opportunity to launch articulation. The following section discusses research that suggests listeners can often prepare their utterances in this way.

#### 1.3.2.3. Evidence for late planning

Listeners could minimize the overlap between production and comprehension processes by preparing a response towards the end of the speaker's turn (i.e., the late-planning hypothesis). One of the main arguments against this proposal is that listeners would not have enough time to prepare their whole response prior to articulation, especially in cases where their response is relatively long or complex, and so could not achieve inter-turn intervals of 200 ms (e.g., Levinson & Torreira, 2015).

However, listeners could still avoid long gaps between utterances and maintain conversational fluency by preparing their response at the same time as launching articulation. Studies of monologue provide extensive evidence that language production can be incremental in this way. For example, Wheeldon and Lahiri (1997) found that utterance initiation times were longer when the first word of the utterance was more phonologically complex. However, initiation times were not influenced by the phonological complexity of later words, suggesting that the time it takes the speaker to produce their utterance is affected by the time it takes them to plan their first word, rather than the time it takes them to produce their complete response. In other words, listeners planned only their first word prior to articulation, while later words were planned while they were speaking (see also Brown-Schmidt & Konopka, 2015). Although these studies have investigated planning during monologue (i.e., without the need to coordinate with another speaker), similar mechanisms may also occur during dialogue.

In one study investigating response preparation in dialogue, Torreira, Bögels, and Levinson (2015) examined the time course of listeners' pre-speech inbreaths, which have been shown to be related to characteristics of the response to be prepared (such as response length; e.g., Fuchs, Petrone, Krivokapić, & Hoole, 2013). When analyzing a corpus of question-answer sequences, they found that inbreaths were more common when the answer was longer rather than shorter. Furthermore, these inbreaths typically occurred around 15 ms after the end of the speaker's question, suggesting that listeners prepared their response towards the end of the speaker's utterance (i.e., consistent with the late-planning hypothesis). However, inbreaths may also be an index of articulation rather than response preparation, and so it is unclear whether these results are consistent with the late-planning hypothesis.

Other studies consistent with late planning have largely used dual-task paradigms, in which participants engage in conversation while simultaneously conducting an unrelated secondary task. These studies assume that performance on a secondary task should decline when participants begin response preparation. Using this method, Boiteau, Malone, Peters, and Almor (2014) had participants complete a visuomotor tracking task while engaging in an unscripted conversation with a confederate. They found that visuomotor tracking performance declined towards the end of the speaker's utterance, and therefore argued that listeners begin response preparation at this moment. Similar results have been found in monologue (Almor, 2008): Speakers are slower to categorize tones played towards the end of their utterances, when they are presumably planning their next turn, than those played at the beginning.

However, these studies did not examine whether listeners prepared their response earlier when they could predict turn content. Sjerps and Meyer (2015) addressed this issue in a further study, in which they instructed participants to carry out a finger-tapping task while listening to pre-recorded descriptions of one of two rows of four pictures. Participants then described the second row. Even though participants knew which pictures they would later have to describe as soon as the speaker produced the first word of their utterance, participants' finger-tapping performance was affected only when the speaker began describing the last picture in their set (around two seconds after they had started speaking), suggesting they delayed (at least some aspects of) response preparation. Together with Boiteau et al. (2014), these studies are consistent with the late-planning hypothesis, and suggest that response preparation and articulation timing are tightly interwoven during turn-taking: Listeners begin preparation only towards the end of the speaker's utterance, when they will soon have the opportunity to launch articulation. In other words, response preparation depends on being able to predict articulation timing, even when content is predictable and listeners can prepare a response before the turn-end.

But although dual-task paradigms might shed some light on the processes of response preparation, it is unclear which stages of preparation this paradigm taps into. Previous research suggests that all stages of response preparation (such as lemma, word form, and phoneme selection; e.g., Cook & Meyer, 2008; V. Ferreira & Pashler, 2002; Roelofs, 2008; Roelofs & Piai, 2011) and possibly articulation and speech monitoring (e.g., Almor, 2008) are cognitively demanding. For instance, Ferreira and Pashler had participants name pictures while discriminating between tones, and found that increasing the time required for lemma selection (by presenting pictures following less constraining sentences) and word-form selection (by presenting pictures with lower frequency names) delayed both picture naming and tone discrimination, suggesting that both these stages are cognitively demanding. However, it is less clear whether phoneme selection requires central processing capacity: Although Ferreira and Pashler found that manipulating the time required for phoneme selection (by presenting pictures with phonologically related distractors) facilitated picture naming but did not affect tone discrimination, Cook and Meyer found that phoneme selection did not interfere with dual-task performance at all. As a result, it is possible that dual-task difficulty only arises towards the end of the speaker's utterance because it is more sensitive to later, rather than earlier, stages of response preparation.

In addition, the secondary tasks (e.g., finger-tapping, visuomotor tracking) involved in these paradigms are nonlinguistic, and often involve processes that are unrelated to the main task. This is of course not the case in conversation, in which participants engage in simultaneous production and comprehension, which are often related: Listeners use production mechanisms to prepare utterances that often complement their comprehension of the speaker's utterance, and thus likely overlap in content (e.g., adjacency pairs; Schegloff, 1996).

The discrepancy in the findings of dual-task studies and others using more naturalistic paradigms (e.g., question-answering; Bögels et al., 2015) may also be attributed to the flexibility of advance planning. In other words, there may be instances in conversation where listeners prepare their response in advance of the speaker's turn-end (i.e., using content prediction to prepare a response is independent from predicting timing), but others where listeners prepare their response only when they know they will soon have the opportunity to articulate (i.e., preparation depends on predicting timing and not on predicting content). The next section discusses this issue in more detail.

#### 1.3.2.4. Evidence for flexible planning

Many authors have stressed that speech planning is flexible (e.g., Swets, Jacovina, & Gerrig, 2013; Konopka, 2012). For example, F. Ferreira and Swets (2002; see also Swets et al., 2013) found that the scope of advance planning (i.e., how much of their response the speaker prepares before speech onset) was influenced by time pressure. Participants produced answers to two digit sums (e.g., 9 + 7 = ?) when time pressure was absent (Experiment 1) or present (Experiment 2). In both experiments, initiation times increased as problem difficulty also increased. However, problem difficulty influenced utterance duration only in Experiment 2, suggesting that speakers simultaneously planned and articulated when they were encouraged to produce their utterance immediately. When there was no pressure, participants made use of more extensive advance planning. Similarly, Wagner, Jescheniak, and Schriefers (2010; Experiment 1) measured planning scope using a PWI task, in which participants were presented with unrelated or semantically related auditory distractors while they produced simple sentences consisting of two nouns (e.g., *the frog is next to the mug*). The authors found that although interference effects for the first noun were similar in size for fast and slow speakers (selected based on their average naming latencies in the unrelated distractor condition), the interference effects on the second noun was larger for the slow than the fast speakers. These results suggest that slow speakers had a tendency to plan further in advance than fast speakers.

The scope of advance planning is also influenced by the content of the previous speaker's turn. Konopka (2012) found that increasing the familiarity of lexical items, by manipulating frequency and recent usage, increased speaker's planning scope from one to two words. This result may have occurred because representations accessed in comprehension were then more accessible during later production, thus facilitating planning. Planning scope is also sensitive to the ease of structural assembly. In their second PWI experiment, Wagner et al. (2010) asked participants to only produce simple sentences (e.g., *the frog is next to the mug*) or to switch between simple and complex sentences (e.g., *the red frog is next to the red mug*). They found that this additional cognitive load eliminated any interference effect for the second noun, regardless of whether speakers were slow or fast. Conversely, Konopka (2012) found that increasing the familiarity of sentence structure (through repetition) increased speaker's planning scope. Together, these studies suggest participants extended the scope of advance planning when structures were repeated and were thus easier to produce.

These results are particularly relevant for conversational turn-taking, since interlocutors in dialogue often align their representations and repeat sentence structures and words previously used by their partner (e.g., Branigan, Pickering, & Cleland, 2000). In a set of studies, Garrod and colleagues (Garrod & Anderson, 1987; Garrod & Clark, 1993; Garrod & Doherty, 1994) found that participants in a maze game tended to converge on descriptions (e.g., participants described positions in the maze as column row indices or as paths between two points) and lexical expressions (e.g., referring to each node in the maze as either *box* or *square*; see also Brennan & Clark, 1996). In addition, Branigan et al. (2000) found that the syntactic structure of participants' picture descriptions was influenced by the structure of a confederate's previous description: When the confederate produced a prepositional object (e.g., *The X verbing the Y to the Z*) or a direct object (e.g., *The X verbing the Z the Y*) description, participants tended to produce the same syntactic form.

As a result, we may expect more advance planning in particular turn-taking exchanges. More specifically, interlocutors may plan more of their response before speech onset when they are aligned with their conversational partner (e.g., when their exchanges involve lexical and structural priming). Although this early planning may be cognitively demanding (see Section 1.3.2), planning may be less cognitively demanding in these instances because the representations the listener requires for production have already been primed during comprehension. Conversely, interlocutors may favor late planning and thus incremental preparation when they are not aligned with their conversational partner because they cannot prepare much of their response in advance of the turn-end.

In sum, studies exploring the scope of advance planning suggest that there are likely some instances in which listeners engage in early planning (because they have more resources available to prepare more of their response before the turn-end) and others in which they engage in late planning (because they have fewer resources available to prepare a response before the turn-end). Given the complexities associated with the scope of advance planning of the content of a response, the studies in this thesis focus only on *yes/no* answers.

#### 1.3.2.5. Conclusion

To sum up, the results of research exploring the time course of response preparation in language production are mixed. Some studies suggest that content and timing predictions are independent, and listeners prepare the content of their response independently from timing response articulation (i.e., the early-planning hypothesis). Other studies, however, suggest that listeners begin preparing the content of their response when they can predict the timing of articulation (i.e., when they know the speaker will soon reach the end of their turn; the late-planning hypothesis). Experiments 1-4 in this thesis evaluate these hypotheses further by examining whether listeners use content predictions to either prepare a response, predict the speaker's turn-end, or both.

## 1.4. Perceptual Learning

Thus far, we have focused on how listeners use prediction to coordinate their utterances during conversational turn-taking. But to successfully prepare an appropriate response and time its articulation, listeners must correctly predict and comprehend the speaker's unfolding utterance. Natural speech tends to vary both within and across talkers, such that the pronunciation of a linguistic unit can vary dramatically depending on who is producing it. Nevertheless, speech comprehension is relatively robust, even under challenging conditions. For example, listeners can successfully comprehend talkers who speak at different rates (e.g., Miller & Liberman, 1979; Tauroza & Allison, 1990), with different accents (e.g., Maye, Aslin, & Tanenhaus, 2008), and in different conversational situations (i.e., formal vs. informal; Krause & Braida, 2004; Liu, Del Rio, Bradlow, & Zeng, 2004).

In fact, listeners can often adapt their comprehension to cope with variations in talker characteristics. For example, Bradlow and Bent (2008) found that listeners were better at comprehending Mandarin-accented spoken sentences after exposure to a Mandarin-accented speaker. This adaptation was talker-specific, however, such that exposing listeners to one speaker during training enhanced intelligibility scores for subsequent test sentences only when they were produced by the same speaker. When listeners were exposed to multiple speakers during training, intelligibility scores were enhanced for sentences produced by novel speakers. Thus, listeners required exposure to multiple Mandarin-accented speakers to learn which characteristics were talker- and accent-specific. Similar learning effects have been observed with more artificial distortion, such as time-compression. For instance, Dilley and Pitt (2010) demonstrated that listeners adapt to variations in speech rate, which influences the perception of subsequent speech such that syllables either are (e.g., *leisure time* is perceived as *leisure or time*) or are not (e.g., *leisure or time* is perceived as *leisure time*; see Section 1.2) comprehended. Additionally, Dupoux and Green (1997) found that comprehension of time-compressed sentences was poor on initial presentation but increased by up to 15% when listeners were exposed to 15-20 training sentences. This effect generalized to speech produced by a different talker and at a different rate, suggesting that it did not simply reflect short-term adaptation.

Together, these studies demonstrate that listeners can adapt to variations in the acoustic input. This adaptation is a form of perceptual learning – “relatively long-lasting changes to an organism’s perceptual system that improve its ability to respond to the environment and are caused by its environment” (Goldstone, 1998, p. 586). In other words, listeners update their processing (or their comprehension) in response to talkers with different characteristics (e.g., speaking rate or accent), which influences later comprehension. But listeners can often predict what they are going to hear before they actually hear it (see Section 1.1). Listeners may be able to use these predictions to guide their interpretation of speech under difficult circumstances.

Experiments 7-9 in this thesis (Chapter 4) investigate this issue in further detail. Thus, the following section discusses studies that have considered the role of top-down knowledge and prediction during perceptual learning.

Much research suggests that top-down (lexical) knowledge plays an important role in perceptual learning. For example, Norris, McQueen, and Cutler (2003) presented participants with 20 words, in which all occurrences of either /f/ or /s/ were replaced with an ambiguous fricative between the two. When listeners subsequently completed a phonetic categorization task, they were more likely to perceive the ambiguous sounds as either /f/ or /s/, depending on which phoneme was replaced during training. Importantly, this perceptual learning effect occurred only for listeners who were exposed to words rather than non-words, suggesting that lexical knowledge plays an important role in perceiving ambiguous fricatives. This interpretation was further confirmed by Leach and Samuel (2007), who found that participants learning novel words over five days increasingly showed a perceptual learning effect for ambiguous fricatives as these novel words became lexicalised.

Using a similar task, McQueen, Cutler, and Norris (2006) extended Norris et al.'s (2003) results and found that learning generalized to words that were not presented during training. After training, participants completed an identity priming task, in which they listened to auditory primes (e.g., *knife*) and then made lexical decisions to visual targets (e.g., *nice*). Participants who heard ambiguous /f/ fricatives during training showed facilitation for /f/-final words in this priming task; those who heard ambiguous /s/ fricatives showed facilitation for /s/-final words. Together, these results suggest that training with ambiguous fricatives benefits recognition of these fricatives in untrained items.

Other studies have found similar learning effects using more artificial distortions, such as noise-vocoding. Noise-vocoding is an acoustic distortion that is created by dividing the speech stream into a number of frequency bands and then applying the amplitude envelope of each frequency range to band-limited noise, thus removing spectral information from the speech signal while still preserving temporal cues (R. V. Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Speech vocoded with more than ten bands is readily intelligible, but decreasing the number of bands reduces intelligibility. In particular, speech vocoded with five to eight bands is around 50% intelligible, while speech vocoded with fewer than four bands is typically difficult to understand (see R. V. Shannon, Fu, & Galvin, 2004).

In one study using this manipulation, Davis, Johnsrude, Hervais-Adelman, Taylor, and McGettigan (2005; see also Jacoby, Allan, Collins, & Larwill, 1988; Remez et al., 1981) presented participants with noise-vocoded sentences and instructed them to type what they heard. After listening to this distorted sentence, participants subsequently heard (Experiment 2) or read (Experiment 3) a clear version of the sentence followed by the distorted version a second time (distorted(D)-clear(C)-distorted(D) condition), or heard the distorted sentence twice before hearing the clear version (DDC condition). The authors found that listeners who knew the identity of the distorted sentence prior to its second presentation (DCD condition) were able to report more words during the first presentation of subsequent vocoded sentences than participants who heard both versions of the distorted sentence before the clear version (DDC condition). In other words, listeners showed more rapid perceptual learning when they knew the identity of the distorted sentence (and had top-down knowledge of its content) prior to its second presentation. Hervais-

Adelman, Davis, Johnsrude, and Carlyon (2008) reported similar results for noise-vocoded words.

This learning effect did not occur when participants were trained with sentences containing non-words (Davis et al., 2005; Experiment 4). However, Hervais-Adelman et al. (2008; Experiment 2) found that participants trained with single non-words showed comparable perceptual learning as participants trained with words during a DCD procedure. This discrepancy may be explained by differences in the memorability of stimuli in the two studies. Specifically, learning may not have occurred for non-word sentences because participants had difficulty maintaining a string of clear non-words in capacity limited phonological memory (cf. Gathercole, Willis, Baddeley, & Emslie, 1994), and so they could not make comparisons between a target representation (or prediction) of the clear stimulus and the distorted versions. When participants were trained with single non-words, however, the phonological representation of the clear form was likely still active when the subsequent distorted version was presented. In other words, perceptual learning can occur as long as participants still have a representation of the clear distorted stimulus. Thus, these studies suggest that although perceptual learning is facilitated by lexical information, it can still occur in the absence of this information (i.e., for non-words) if listeners can retain predictions regarding the form of the distorted speech in memory.

Studies demonstrating effects of top-down knowledge during learning are consistent with interactive accounts of speech perception, such as TRACE (e.g., McClelland & Elman, 1986), which claim that higher-level lexical representations can immediately influence lower-level auditory processes through feedback connections. In other words, top-down lexical information is used to fine tune bottom-up pre-lexical processing to ensure utterances are comprehended correctly. These models contrast with accounts that suggest speech comprehension is strictly feedforward (e.g., Merge; Norris, McQueen, & Cutler, 2000), such that early bottom-up acoustic processes cannot be influenced by top-down processes until a later decision stage, at which a final interpretation (e.g., a word report judgement) is formed.

Although it is beyond the scope of this thesis to review the evidence extensively, there is considerable debate concerning which of these accounts is likely to be correct. For example, some studies demonstrate that listeners are faster at identifying phonemes in words than non-words (e.g., Cutler, Mehler, Norris, & Segui, 1987), suggesting that top-down information influences perception. However, this pattern may also be explained by a strictly feedforward model. Specifically, in their Merge model, Norris et al. argue that pre-lexical representations activate their corresponding lexical items, which in turn activate the relevant decision nodes. As a result, decision nodes that have received activation from both the pre-lexical and lexical levels (i.e., phonemes presented in words) will be activated to a greater extent than decision nodes receiving only pre-lexical activation (i.e., phonemes presented in non-words), which leads to faster identification of phonemes embedded in words than those in non-words.

Norris et al. (2000) also argued that top-down effects during perceptual learning do not necessarily suggest that the processes of speech perception are interactive. Specifically, they distinguish between two types of feedback: (i) Feedback for online perception, in which higher-level lexical knowledge immediately constrains processing at the pre-lexical levels, and (ii) feedback for learning, in which higher-level knowledge is used to permanently adjust pre-lexical representations, so that all future utterances (including novel items) are interpreted using these representations (e.g., interpreting ambiguous fricatives as /f/, even in words that were not heard during training). Thus, it may be possible to explain lexical effects during perceptual learning without necessarily assuming that language processing is interactive.

But how does top-down knowledge aid perceptual learning? Research demonstrating lexical effects during learning can be interpreted in line with a predictive coding account (e.g., Arnal & Giraud, 2012), in which sensory representations are used to predict the most likely upcoming events. These predictions are then compared with incoming information and the difference between the two (the prediction error) is carried forward to alter future processing. Thus, listeners presented with a clear version of the stimulus prior to distortion (i.e., in the DCD training condition in Davis et al.'s (2005) study) use this representation to predict the form of the distorted input. Any difference between the two yields an error signal, which is used to adjust later representations so that they more closely match the incoming speech input.

Sohoglu, Peelle, Carlyon, and Davis (2012; see also Sohoglu & Davis, 2016) found results consistent with this account. They manipulated prior knowledge of distorted speech, so that participants were presented with matching (text that matched the distorted word), mismatching (text that matched a different distorted word), or neutral (a string of 'x' characters) written text prior to the presentation of words that were noise-vocoded using two, four, or eight bands. Behavioral results showed that participants gave higher clarity ratings (on a scale of 1-8), which are strongly related to word report scores (Davis & Johnsrude, 2003), to (i) noise-vocoded words preceded by matching rather than mismatching or neutral text, and (ii) words vocoded with more bands. Additionally, concurrent MEG and EEG recordings showed reduced activity in the inferior frontal gyrus when distorted speech was preceded by matching rather than mismatching or neutral text. Such reduced activity is associated with the processing of speech content (e.g., Scott & Johnsrude, 2003) and is thought to occur because listeners use prior knowledge to predict incoming speech input, thus reducing prediction error. Conversely, activity is increased when distorted speech is preceded by mismatching text because prediction error is also increased. This effect occurred before reduced activity in the superior temporal gyrus, which is associated with lower-level sensory processing, providing further evidence for top-down processing.

Similar results were reported by Blank and Davis (2016), who found that matching text and increasing sensory detail (speech vocoded with twelve bands compared to four) both improved word report scores and reduced BOLD signals in the lateral temporal lobe. But these two factors also interacted, such that sensory detail increased the amount of information represented in superior temporal multivoxel patterns (which measure how much information about the phonetic form of speech is contained in functional Magnetic Resonance Imaging (fMRI) activation patterns) when prior knowledge was uninformative; when prior knowledge was informative, however, increased sensory detail reduced the amount of information represented in multivoxel patterns.

However, such findings may also be attributed to ease of integration. In other words, listeners do not use prediction to guide perceptual learning. Instead, faciliatory effects from written or auditory presentation of the clear stimulus prior to distortion (relative to conditions in which the clear stimulus is presented after distortion or in which the stimulus does not match the distorted text) could be attributed to increased ease of integrating the lexical representations of distorted speech into unfolding representations (see Kutas et al., 2011, for a review of prediction vs. integration accounts). For example, distorted words that match a previous clear presentation are more plausible than distorted words that do not, and research suggests that greater plausibility results in faciliatory effects, such as faster reading times (e.g., Rayner et al., 2004). Thus, these experiments do not allow us to tease apart perceptual learning effects reflecting ease of integration and prediction error, as the faciliatory effect could be attributed to either or both.

The problem of distinguishing between prediction and integration also affects a number of other studies. For example, Signoret, Johnsrude, Classon, and Rudner (2018) presented participants with noise-vocoded sentences that were either semantically coherent, and thus constrained the number of potential continuations (e.g., *Her daughter was too young for the disco*), or semantically incoherent, and did not provide any information about the content of the speaker's forthcoming words (e.g., *Her hockey was too tight to walk on cotton*). The authors found that clarity ratings (on a scale of 1-7) were higher when these sentences were (i) semantically coherent rather than incoherent, and (ii) preceded by matching rather than mismatching written text. Based on these results, Signoret et al. concluded that both semantic and phonological form-based predictions aid perceptual clarity.

In a similar study, Davis, Ford, Kherif, and Johnsrude (2011) found higher word report scores for semantically coherent (around 40%) than incoherent (around

20%) distorted sentences, suggesting that participants can predict the form of distorted speech from the speaker's preceding words without necessarily hearing a clear repetition of these words prior to the vocoded version. Additionally, the magnitude of fMRI activity in frontal and temporal regions depended on sentence clarity and coherence, such that activity was high for degraded speech (regardless of whether it was semantically coherent or not) and clear semantically anomalous speech, but low for clear semantically coherent sentences. However, the timing of this activity occurred earlier in the temporal (lower-level auditory) than the frontal (semantic) regions. This finding is inconsistent with top-down accounts, which predict the opposite pattern of activity.

However, the findings of both of these studies could still reflect ease of integration: Predictable words are likely more plausible continuations than less predictable words, thus leading to enhanced clarity ratings and word report scores for the semantically coherent than incoherent sentences. Furthermore, neither of these studies assessed perceptual learning, and so it is unclear whether sentence constraint enhances learning in the same way as stimulus repetition (e.g., Davis et al., 2005). Although semantic coherence induces perceptual pop-out, meaning that it is easier for participants to recognize the words in distorted sentences (e.g., Giraud et al., 2004), it may not make it easier to understand novel distorted stimuli. In fact, such perceptual pop-out could reflect response bias: When listening to semantically coherent sentences, it may be easier to guess subsequent words which may make it easier to understand those words when they are distorted.

In sum, experiments showing faciliatory effects of meaningful feedback on perceptual learning have tended to conflate manipulations of predictability and plausibility, and so it is unclear whether perceptual learning reflects prediction or ease of integration. Using a novel manipulation, Experiments 7- 9 (Chapter 4) in this thesis investigate this issue further by independently manipulating the predictability and plausibility of noise-vocoded speech to determine which of these factors aid perceptual learning.

## 1.5. Summary

To summarize, much evidence suggests that listeners predict the content and the timing of upcoming language during comprehension. Although some research suggests that these predictions play a role during conversational dialogue, we pointed out several unanswered questions from past studies. This thesis aims to fill the gaps of existing findings, focusing on two mechanisms that arguably play an important role during conversational dialogue: (1) conversational turn-taking and (2) comprehending utterances in difficult circumstances (perceptual learning).

During conversational turn-taking, there is often little gap between interlocutors' utterances, and thus listeners must ensure that they prepare a response and appropriately time its articulation (i.e., so they do not overlap with the previous speaker). But it is unclear what role prediction plays in these processes, given that listeners must manage the cognitive demands of preparing a response and timing articulation while simultaneously allocating resources to comprehending the speaker's incoming turn.

Experiments 1-4 (Chapter 2) investigated this issue by asking whether listeners use predictions of turn content (i.e., predictions of what the speaker is going to say) to (i) predict the end of the speaker's question, or (ii) prepare a verbal response. To assess these two mechanisms, participants either pressed a button when they thought the speaker was about to finish (Experiments 1 and 3) or verbally answered with either *yes* or *no* (Experiments 2 and 4). Since it is unclear whether listeners can predict turn-endings using predictions of turn length (i.e., syntactic structure), these studies also considered the role that length predictability plays during both of these mechanisms.

Turn-end prediction is unlikely to be the only mechanism used for timing response articulation, and so Experiments 5 and 6 (Chapter 3) considered other processes involved in response timing. Specifically, research has demonstrated that listeners can make timing predictions based on speech rate entrainment during comprehension (see Section 1.2), but very little has considered whether listeners use these timing predictions to time response articulation during dialogue (Section 1.3.2). Experiments 5 and 6 investigated this issue by manipulating the speech rate of utterances during a *yes/no* question-answering task.

Finally, response preparation and articulation rely on successfully comprehending the speaker's turn. Thus, the final experiments in this thesis (Experiments 7-9; Chapter 4) investigated whether listeners generate detailed perceptual predictions of upcoming language, which may help them understand speech under difficult circumstances, such as when speech is distorted.

# 2. Study 1 Experiments 1-4: The role of content predictions in response preparation and turn-end prediction[4]

## 2.1. Introduction

Speaking and listening to speech are both extremely complex processes. Yet, during conversation interlocutors are able to switch from one to the other exactly when they need to. In fact, speakers rarely overlap extensively, and the gap between their turns typically averages 200 ms (Stivers et al., 2009). To achieve such coordination, listeners must prepare their own response and articulate it at the appropriate moment. But how do they do so?

Current theories agree that interlocutors achieve such coordination in part by predicting the content of the speaker's incoming turn (i.e., what the speaker is likely to say next; e.g., Bögels & Levinson, 2017; Garrod & Pickering, 2015). Indeed, we know that comprehenders can predict upcoming language at different linguistic levels, including semantic, syntactic, and form-related information (e.g., Altmann & Kamide, 1999; Van Berkum et al., 2005). However, it is currently unclear how these content predictions aid successful turn-taking.

Such predictions may ease processing of the incoming turn, allowing listeners to prepare an appropriate response (e.g., one which is semantically and syntactically

[4] Experiment 1 in this study was designed and carried out by the author in collaboration with Abigail Crossley, who submitted this work as part of her undergraduate dissertation for a degree in Psychology at the University of Edinburgh. This chapter is based on a pre-proofed manuscript published in Cognition (Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition, 175,* 77-95.).

appropriate) in good time, and thus respond earlier. But on its own, early preparation may not be sufficient for smooth turn-taking: Listeners must also articulate their response at the appropriate moment, so they do not overlap with the previous speaker nor leave a long gap. Content predictions may help listeners predict when the speaker's turn will end (see Corps, Gambi, & Pickering, 2018), so they can time their responses more precisely (i.e., clustered closer to the turn-end).

In principle, content predictions might support smooth turn-taking both by facilitating earlier response preparation and by allowing more precise turn-end prediction. Crucially, however, it is currently unclear how the process of determining what to say relates to the process of determining when to speak. One possibility is that listeners use content predictions to prepare a response early, hold this response in an articulatory buffer, and then launch articulation reactively when the speaker displays turn-final cues (e.g., drawl on the final syllable; Duncan, 1972). We term this the *early-planning hypothesis* (e.g., Levinson & Torreira, 2015), as it proposes that listeners determine what to say early, separately from determining when to say it. According to this hypothesis, content predictability facilitates turn-taking because listeners can prepare a response earlier when the content of the speaker's turn is more rather than less predictable. This account predicts that there is no role for prediction of the speaker's turn end because listeners use turn-final cues to determine when to speak, and so content predictability should only benefit the process of determining what to say and not the process of determining when to say it.

But turn-final cues are far from perfect predictors of a turn change (e.g., Gravano & Hirschberg, 2011). In addition, using production processes to prepare and buffer a response is cognitively demanding and may interfere with the listener's ability to comprehend the speaker's unfolding utterance. Importantly, listeners could avoid such interference by beginning preparation only when they believe that they will soon have the opportunity to articulate their response (i.e., late in the turn; Sjerps & Meyer, 2015). According to this *late-planning hypothesis*, listeners use content predictions to predict the speaker's turn-end and only begin response preparation close to this moment (cf. Bögels & Levinson, 2017). If this is the case, then listeners should be more precise at predicting the speaker's turn-end when content is more rather than less predictable.

Note that although we present two opposing accounts in line with the literature, they are not necessarily mutually exclusive. The two mechanisms could work in parallel to some extent (see Bögels & Levinson, 2017). For example, listeners could use content prediction to prepare a response early and also to predict the speaker's turn-end in order to better time response articulation, in a way that would combine elements of both the early planning and the late planning account. Conversely, listeners may prepare late and also use turn-final cues (rather than turn-end prediction) to time articulation. However, it is an empirical question whether predictability affects only response preparation (early-planning), only turn-end prediction (late-planning), or indeed both.

To explore the role of predictability during turn-taking, we manipulated the content predictability of simple *yes-no* questions in two pairs of experiments, using two paradigms designed to capture different aspects of the turn-taking process. To isolate turn-end prediction, we first used a button-press task, in which listeners pressed a button as soon as they expected the speaker to reach the end of their turn (i.e., they were encouraged to predict this moment; De Ruiter et al., 2006). Since this paradigm encourages participants to precisely time their response, we analyzed absolute response precision (i.e., how close participants responded to the speaker's turn-end). While the early-planning hypothesis does not predict any difference in precision between predictable and unpredictable questions (because it assumes no role for turn-end prediction), the late-planning hypothesis predicts that listeners should be more precise (i.e., their responses should cluster closer to the speaker's turn-end) when they can predict question content than when they cannot.

To further explore the role of content predictability, we conducted two additional experiments using a question-answering task, which we assume captures response preparation in addition to turn-end prediction. Accordingly, we analyzed not only the precision of participants' responses (as in the button-press task), but also the signed response times (i.e., how early participants responded). Precision and response times are of course related measures but, crucially, can influence response precision in different ways: If participants are slower to respond, their responses can become either less precise (if they occur after the end of the speaker's turn) or more precise (if they occur before the end of the speaker's turn). Moreover, changes in precision can occur independently of changes in response time (e.g., if the spread of responses increases without changes to the mean response time).

Thus, it is necessary to analyse both measures to determine whether content predictability affects precision (i.e., as predicted by the late-planning hypothesis) and whether it affects response timing (i.e., as predicted by the early-planning hypothesis). Early-planning proposes that listeners should respond earlier when they can predict question content than when they cannot (because content prediction helps listeners prepare earlier), but does not predict any difference in precision between predictable and unpredictable questions (because articulation is timed based on a different mechanism, namely reaction to turn-final cues). In contrast, the late-planning hypothesis proposes that responses should be more precise for predictable than unpredictable questions (because prediction helps listeners determine the turn-end more accurately), but does not predict any difference in signed response times between predictable and unpredictable questions (because listeners always begin preparation close to the turn end anyway).

We used the same items in both tasks to ensure comparability between the experiments. In the rest of the Introduction, we discuss evidence for and against both accounts, before describing the current study and formulating our predictions in more detail. We also distinguish two versions of the late-planning account that differ in what information they assume is used for turn-end prediction.

### 2.1.1. Evidence for early planning

Some research suggests that listeners prepare their own turns as early as possible. For example, in a question-answering task Bögels et al. (2015) found that participants responded earlier and showed activation in brain areas involved in speech production (e.g., Indefrey & Levelt, 2004) and motor response preparation (e.g., Bablioni et al., 1999) when the information (here, *007*) necessary for response preparation was available early in the turn (e.g., *Which character, also known as 007, appears in the famous movies*?) rather than late (e.g., *Which character from the famous movies is also called 007*?). These results suggest participants prepared their response further in advance when the critical information was available early rather than late. Importantly, they did so even though the question could have continued in a number of different ways (e.g., *appeared in Skyfall?, was recently played by Daniel Craig?*), meaning they could not necessarily predict the turn-end.

Barthel et al. (2016) provided further support for the early-planning account using a list-completion task, in which participants completed a confederate's pre-recorded utterances. Participants had to name any on-screen objects that the confederate had not already named, and so they could (in principle) prepare their response as soon as the confederate began uttering the last object name. The authors also manipulated whether participants could predict that the speaker's turn would end with a turn-final verb. Both eye-movements and response latencies suggested that participants planned their response as soon as possible. However, neither of these measures were influenced by the predictability of the speaker's turn-end, suggesting that listeners did not use such predictions to time response articulation. Participants may instead have launched articulation using turn-final cues (see Barthel et al., 2017).

### 2.1.2. Problems with early planning

Although the evidence in Section 2.1.1 supports the early-planning hypothesis, this account faces two unresolved issues. First, it is unclear whether turn-final cues can explain all turn-taking behaviour. In a corpus study of dyadic interactions, Gravano and Hirschberg (2011) assessed the role of seven turn-final cues (e.g., lengthening of the final word) and found that these cues were significantly more likely to occur in stretches of speech preceding speaker changes than in those preceding a continuation of the current speaker's turn. However, listeners were only 65% likely to take a turn when all seven cues were present. Although one of the cues considered by the authors was whether the turn was semantically and/or syntactically complete, they did not explore the role of content predictability, thus leaving open the possibility that other content-based mechanisms (such as turn-end prediction) are also at play.

Second, if addressees prepare their response as soon as possible, then production and comprehension processes must overlap. Since these processes recruit overlapping neural circuits (e.g., Segaert et al., 2012) and most likely share resources, using production mechanisms to prepare and buffer a response in advance of the turn-end should be cognitively demanding and may interfere with the concurrent process of comprehending the speaker's turn. Indeed, previous research suggests all stages of preparation (e.g., lemma, word form, and phoneme selection; Cook & Meyer, 2008) require central processing capacity.

Crucially, listeners could avoid such interference by preparing a response only when they are sure the speaker is about to finish (i.e., late-planning hypothesis). Sjerps and Meyer (2015; see also Boiteau et al., 2014) found results consistent with this account using a dual-task paradigm, in which participants completed a finger-tapping task while listening to pre-recorded picture descriptions. Even though participants knew which pictures they would later have to describe as soon as the speaker produced the first word of their utterance, participants' finger-tapping performance was affected only when the speaker began describing the last picture in their set (around two seconds after they had started speaking), suggesting that participants delayed response preparation. Contrary to Bögels et al. (2015), these studies support the late-planning hypothesis and suggest that listeners begin preparation towards the end of the speaker's turn.

### 2.1.3. Turn-end prediction: Dissociating content from length predictability

For the late-planning hypothesis to be correct, listeners must be able to determine when the speaker's turn will end so they can begin response preparation at the appropriate moment. However, it is still largely unclear how listeners predict turn-ends.

So far in our discussion of the late-planning hypothesis, we have assumed that listeners use content predictions (i.e., lexico-semantic properties of upcoming words) to determine the speaker's turn-end. However, listeners may also predict the length of a turn by separately estimating the number of words until turn-end (e.g., Magyari & De Ruiter, 2012). Indeed, utterances are often predictable in length but unpredictable in content. To illustrate, the sentence fragment *Most people have two…* can be completed with many single words (e.g., *cars, dogs, siblings*), which overlap very little in their content. Conversely, utterances can be unpredictable in length but predictable in content. For example, the sentence fragment *The Titanic sank after…* can be completed with *it hit an iceberg, hitting an iceberg,* or *crashing*, which differ in length but overlap in content. Thus, listeners could predict a speaker's turn-end by predicting either its lexico-semantic content or its length (in number of words). Of course, being able to predict the length of the turn in number of words may not be sufficient to predict the turn-end accurately, as words differ in duration (e.g., number of syllables). However, such predictions would greatly constrain estimates of turn duration.

Given this distinction, one version of the late-planning hypothesis (*the length-prediction hypothesis*) proposes that turn-end prediction should be more precise when length is predictable rather than unpredictable, regardless of content predictability. For example, Magyari and De Ruiter (2012) found that turns that participants expected to be completed with more words (even though they could not predict the exact words) were those that elicited later button-press responses, suggesting that listeners can predict turn-ends by predicting the number of words the speaker will use.

The length-prediction hypothesis contrasts with a second version of the late-planning hypothesis, which we term the *content-prediction hypothesis*. This version maintains that length predictions are possible only when content is predictable. When content is unpredictable, listeners should not be able to predict how many words will follow. For example, Magyari, et al. (2014) found that participants responded 70 ms before the end of predictable turns but 139 ms after the end of unpredictable turns. Together with concurrent EEG recordings, these results suggest that listeners used turn content to predict the speaker's turn-end.

However, previous studies have not manipulated length predictability independently from content predictability. In this study, we thus investigated whether participants predicted the length (in number of words) of the speaker's question, and whether they did so independently of predictions of content. To do so, we crossed our manipulation of content predictability (predictable vs. unpredictable; i.e., whether participants could predict the lexico-semantic content of upcoming words) with a manipulation of length predictability (single vs. varied; i.e., whether participants expected a single word completion or had no clear expectation about the number of words that would follow; see Table 1 for example stimuli) of simple questions.

Table 1. Example materials and possible completions for each of the four stimuli conditions in Experiments 1 and 2.

| Content Predictability | Length Predictability | Example Question Fragment | Possible Completions |
|---|---|---|---|
| Predictable | Single | Are dogs your favorite…? | animal |
| | Varied | Did The Titanic sink after…? | it hit an iceberg/hitting an iceberg/crashing |
| Unpredictable | Single | Do you enjoy going to the…? | supermarket/dentist/beach |
| | Varied | Do most students finish their…? | dinner/studies after four years/exams on time |

Note that the early-planning hypothesis is not concerned with the distinction between content and length prediction, as it assumes no role for turn-end prediction. However, both versions of the late-planning account predict that listeners' button-press (Experiments 1 and 3) and question-answering (Experiments 2 and 4) responses should be more precise when content is predictable than when it is not. The content-prediction hypothesis predicts an interaction between content and length predictability, such that listeners should be more precise when length is predictable than when it is not, but only when content is also predictable. In contrast, the length-prediction hypothesis proposes that listeners should be more precise when length is predictable rather than unpredictable, regardless of content predictability. Finally, recall that since the early-planning hypothesis assumes that turn-end prediction does not play a role, it does not predict any effects of either content or length predictability on the precision of responses in any of the experiments.

### 2.1.4. Overview of Experiments

In sum, we do not know how response preparation and articulation are interwoven during conversational turn-taking. Listeners may achieve such coordination by preparing a response early and launching articulation only after a turn-final cue (the early-planning hypothesis; Levinson & Torreira, 2015). Alternatively, they may begin preparation only when they know that the speaker is soon going to reach the end of their turn (the late-planning hypothesis; Sjerps & Meyer, 2015) and they may predict the turn-end either by predicting turn content (content-prediction hypothesis) or by predicting both turn content and turn length (length-prediction hypothesis).

To test these accounts, we conducted two pairs of experiments using button-press (Experiments 1 and 3) and question-answering tasks (Experiment 2 and 4). In Experiments 1 and 2, we manipulated both the content (predictable vs. unpredictable) and length predictability (single vs. varied) of questions, to create four conditions. Experiments 3 and 4 were modelled on Experiments 1 and 2, respectively, but included only three of the four conditions (predictable single, unpredictable single, unpredictable varied) which are sufficient to tease apart the content prediction and the length prediction hypotheses.

In the first pair of experiments, we strengthened participants' expectations about question length by having questions that were unpredictable in length end with a varied number of words (two or more); questions whose length was predictable always ended with a single word. Since this approach made it difficult to compare content predictability across the single and varied conditions, in the second pair of experiments we selected single word completions for all questions (i.e., both those that were unpredictable and those that were predictable in length). Importantly, we found the same pattern of results across both pairs of experiments, suggesting that the length of completions chosen for the varied length conditions did not affect the results.

We analyzed both the response times (i.e., the signed deviation of listeners' responses from the turn-end) and absolute precision (i.e., how clustered around zero participants' response were) of responses in all experiments. However, precision is the most relevant measure for the button-press task, as participants are asked to respond exactly when they think the speaker will reach the end of their turn. In contrast, both response times and precision are relevant for the question-answering task, because this task captures both response preparation and turn-end prediction.

The early-planning account argues that listeners use prediction to prepare a response early, and so they should produce their verbal responses earlier when content is predictable rather than unpredictable. Since this account assumes no role for turn-end prediction, it makes no predictions regarding the precision of participants' responses. In contrast, the late-planning account argues that listeners use prediction to determine the speaker's turn-end, and so their responses should be more precise when the content (and possibly the length) of the speaker's turn is predictable rather than unpredictable. Since this account assumes no role for early preparation, it makes no predictions for effects on response times (see Table 2 for a summary of predictions).

Table 2. Summary of predictions made by the accounts for the button-pressing task, which taps into turn-end prediction (Experiments 1 and 3), and the question-answering task, which taps into turn-end prediction and response preparation (Experiments 2 and 4).

| Measure[a] | Button-press task | Question-answering task |
|---|---|---|
| Early-planning hypothesis | | |
| Response times | No predictions about the effects of content and length predictability on response times during button-pressing. | Content predictability: earlier responses for predictable than unpredictable questions.<br>No predictions about the effects of length predictability during question-answering. |
| Late-planning hypothesis (content-prediction) | | |
| Precision | Content predictability: more precise when content is predictable than unpredictable<br>Length predictability: no main effect on precision.<br>Content*Length predictability: more precise when length is predictable than when it is not, but only when content is predictable. | |
| Late-planning hypothesis (length-prediction) | | |
| | Content predictability: more precise when content is predictable than unpredictable.<br>Length predictability: more precise when length is predictable than unpredictable. | |

[a] Note that the early-planning hypothesis makes different predictions for button-pressing and question-answering, while the late-planning hypotheses make the same predictions for button-pressing and question-answering.

## 2.2. Experiment 1

Experiment 1 used a button-pressing task with four conditions. Stimuli in the single conditions were completed with a single word by the large majority of participants in a cloze pre-test, and were therefore predictable in length. Crucially, this word (in bold in the following examples) was either the same across participants (predictable single; e.g. *Are dogs your favorite* ***animal****?*), so that both content and length were predictable, or different (unpredictable single; e.g., *Do you enjoy going to the* ***supermarket****?*), so that length was predictable but content was not. Stimuli in the varied conditions were followed by completions that varied in length (i.e., their length was not predictable) and either did overlap in content (predictable varied; *Did The Titanic sink after* ***it hit an iceberg****?*), so that content was predictable while length was not, or did not overlap in content (unpredictable varied; *Do most students finish their* ***exams on time****?*), so that neither content nor length were predictable.

### 2.2.1. Method

#### *2.2.1.1. Participants*

Thirty native English speakers (3 males; *M*age = 20.23 years) at the University of Edinburgh participated in exchange for partial course credit or £4. Participants had no known speaking, reading, or hearing impairments.

#### *2.2.1.2. Materials*

We selected 116 questions (29 for each condition) using a norming task, in which 33 further participants from the same population (8 males; *M*age = 20.67) were presented with 160 question fragments and were instructed to "complete with the words or words that you think are most likely to follow the preceding context of the question" (i.e., we used a cloze task; Taylor, 1953).

We assessed length predictability by calculating the sample variance of the length (in number of words) of the completions for each fragment. In the single conditions, participants completed fragments with one word at least 90% of the time and so the length (i.e., a single word completion) was predictable. In contrast, different participants completed fragments in the varied conditions with different numbers of words (higher variance; $p < .001$, see Table 3), and so length was unpredictable. For these fragments, no more than 20% of pre-test participants provided a completion of the same length as the selected multiword completion (which was between two and eight words; $M = 3.22$).

Table 3. The means (and standard deviations) of our measures of content predictability, length predictability, difficulty, plausibility, and duration (ms) for stimuli in Experiments 1 and 2. The final row provides the number of utterances characterized by a pitch downstep in each condition.

| | Predictable | Predictable | Unpredictable | Unpredictable |
|---|---|---|---|---|
| | Single | Varied | Single | Varied |
| Average Completion Length Variance | 0.02 (0.04) | 1.18 (0.82) | 0.11 (0.09) | 0.95 (0.44) |
| Completion Length Cloze[a] | 99% (3%) | 19% (14%) | 92% (8%) | 18% (15%) |
| Question Fragment LSA[b] | .91 (.11) | .71 (.14) | .37 (.12) | .35 (.11) |
| Completion LSA[c] | .95 (.06) | .68 (.19) | .16 (.08) | .23 (.12) |
| Completion Content Cloze[d] | 93% (8%) | - | 4% (2%) | - |
| Question Fragment Entropy[e] | 0.35 (0.36) | - | 3.01 (0.63) | - |
| Question Difficulty[f] | 6.22 (0.48) | 6.11 (0.35) | 6.17 (0.42) | 6.24 (0.40) |
| Question Plausibility[g] | 6.64 (0.35) | 6.45 (0.27) | 6.52 (0.40) | 6.48 (0.39) |
| Question Duration (ms) | 2398 (646) | 2996 (620) | 1932 (452) | 2542 (597) |
| Downstepped utterances | 29/29 | 27/29 | 26/29 | 27/29 |

[a] Percentage of participants who provided the word length of the selected completion used in the main experiment (a single word in the single conditions; multiple words in the varied conditions) as a continuation in the cloze task.

[b] Average over all completion comparisons for that particular fragment.

[c] Average over comparisons between the selected completion and all other completions.

[d] Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

[e] Entropy of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.

[f] Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, while 7 indicated that the question was very plausible/easy to answer.

We assessed content predictability using three different measures. First we calculated cloze probability (Taylor, 1953), which is the percentage of participants who provided a particular completion. We also computed Shannon entropy (i.e., $-\Sigma p_i \mathit{log}_2(p_i)$, where $p_i$ is the proportion of times each completion occurs for a given fragment; C. E. Shannon, 1948). Entropy is low (a minimum of 0) when completions are similar across participants, and high (a maximum of 5.04 when each of the 33 participants in the pre-test provided a different response) when responses are different. Note that both of these measures can only be computed for stimuli in the single conditions, as completions in the varied condition may differ verbatim while having similar content (e.g., *it hit an iceberg* vs. *hitting an iceberg*). Stimuli in the predictable single condition had higher cloze probability ($p < .001$; see Table 3) and lower entropy ($p < .001$) than those in the unpredictable single condition ($p < .001$; see Table 3).

Finally, we computed Latent Semantic Analysis (LSA; Deerwester, Dumais, Furnas, Landauer, & Harsman, 1990) matrix comparisons using the general reading corpus. LSA determines the semantic similarity of words and phrases by calculating the extent to which they occur in the same context, and ranges from 1 (completions are identical) to -1 (completions are completely different). Importantly, it can be used to assess the similarity of completions that differ in number of words.

Using these LSA comparisons, we first calculated the content predictability of each fragment by averaging over the LSA scores for all pairwise comparisons between completions. Stimuli in the predictable content condition had higher fragment LSA than those in the unpredictable content conditions (*p* < .001; see Table 3). We also calculated the LSA value of each completion by averaging over the LSA scores for all comparisons between the chosen completion and every other completion to the same fragment. Completion LSA was higher in predictable than unpredictable conditions (*p* < .001).

The four conditions were matched for average difficulty and plausibility (all *p*s > .07; see Table 3) using data collected in a second pre-test, in which 15 new native English speakers (2 males; *M*age = 19.40) rated (i) how difficult they would find it to answer the question if asked, and (ii) whether the question made sense. Both ratings were made on a scale of 1 (very implausible/difficult to answer) to 7 (very plausible/easy to answer).

All questions were recorded by a native English male speaker, who was instructed to read the utterances as though "you are asking a question and expecting a response". Recordings were between 1317 and 5194 ms in duration (see Table 3). Utterances in the varied conditions were longer than those in the single conditions (*p* < .001), and those in the predictable condition were also longer than those in the unpredictable condition (*p* < .001; we return to this issue in the Results). All our questions had falling boundary tones, and 109 (see Table 3) were characterized by a pitch downstep, which occurs when the pitch of each syllable is lower than the previous syllable (Beckman & Pierrehumbert, 1986). Both judgments were validated by a second rater, who listened to 25% of the utterances (Cohen's kappa = 1, for both ratings).

#### *2.2.1.3. Procedure*

The experiment was controlled using E-Prime (version 2.0). Participants pressed a button to start audio playback of the question. A fixation cross (+) appeared 500 ms before question onset, and the screen turned red as audio playback began. Using a translation of the instructions used by De Ruiter et al. (2006), participants were told: "Press the button (using your dominant hand) when you believe the question will end. Do not wait until the speaker has finished the question and stopped speaking. Instead, you should press the button as soon as you expect the speaker to finish". Thus, they were encouraged to predict the turn-end, rather than simply wait for the speaker to reach the end of his utterance. Participants responded by pressing the middle button of a SR-box and audio playback stopped as soon as a response was recorded (as in De Ruiter et al., 2006).

Participants completed ten initial practice trials to familiarize themselves with the experimental procedure. The 116 stimuli were individually randomized, and participants were given the opportunity to take a break every 29 items.

### 2.2.2. Data Analysis

Precision analyses are most relevant for this experiment, because the button-press task encourages participants to accurately predict the turn end. The late-planning hypothesis predicts effects of content predictability (and possibly length predictability, depending on whether participants make separate content and length predictions) on the precision of participants' button-press responses, whereas the early-planning hypothesis does not predict any differences in precision. In addition, and for comparison with Experiment 2, we also analyzed signed response times. Response times were defined with respect to question offset, and were negative when participants responded before the end of the speaker's question and positive when they responded after the end. We replaced 23 (0.66%) response times falling at least 2.5 standard deviations above the by-participant mean and 96 (2.76%) response times below the by-participant mean with the respective cut-off value. Note that, throughout our analyses, the results were the same regardless of whether or not responses were replaced with cut-off values. We evaluated the effects of content and length predictability on response times with linear mixed effects models (LMM; Baayen, Davidson, & Bates, 2008) using the *lmer* function of the *lme4* package (version 1.1-12; Bates, Maechler, Bolker, & Walker, 2015) in RStudio (version 0.99.896) with a Gaussian link function.

Precision was defined as the absolute value of response time. Before taking the absolute value, we first standardized response time to have a mean of zero, so that we could assume a half-normal distribution or, equivalently (Leone, Nelson, & Nottingham, 1961), a normal distribution truncated at zero. Given that the distribution of response precision is truncated at the lower boundary of zero, the distributional assumptions of lmer are not met. Therefore, we used Bayesian mixed effects models (BMM) as implemented in the *brms* package (version 1.6.1; Bürkner, 2017). We initially fitted models using a normal distribution truncated at zero.

However, such models did not converge, so we modelled our data using three other distribution families: the log-normal, the gamma, and the Weibull distribution (e.g., Pinder, Wiener, & Smith, 1978). In all cases, the Weibull was a better fit than either the log-normal or the gamma (assessed using LOO comparisons), and so we report parameters and credible intervals from models fitted using a Weibull distribution. We ran 4 chains per model, each for 1600 iterations, with a burn-in period of 800, and initial parameter values set to zero. All of the reported models converged with no divergent transitions (all $\widehat{R}$ values ≤ 1.1); the number of effective samples for each estimate is reported in the Appendix.

Although the parameterization of the Weibull distribution implemented in *brms* is based on a scale and a shape parameter, we report and discuss only scale parameters; shape is most often used to model failure or mortality rates, which is not relevant to response precision (although full models are reported in the Appendix). The scale parameter, on the other hand, quantifies the spread of the distribution and is thus informative of the degree of precision in participants' responses. Note that scale parameters were fitted on the log scale (reported in the Appendix), but we report exponentiated estimates in the Results section as they are easier to interpret: The larger the exponentiated value of the scale parameter, the more spread out the probability mass of the distribution. All distributions were fitted using default *brms* priors.

In all instances, we fitted models using the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), except that correlations among random effects were fixed to zero to aid convergence (see Matuschek, Kliegel, Vasishth, Baayen, & Bates, 2017). We fitted the full model where response times or precision was predicted by Content predictability (reference level: unpredictable vs. predictable), Length predictability (reference level: varied vs. single), and their interaction. These predictors were contrast coded (-0.5, 0.5) and centered. We also included Question Duration in our analyses (which was centered), since previous research suggests that longer turns tend to elicit earlier button-press responses (e.g., De Ruiter et al., 2006). To aid convergence, this predictor was included only as a main effect.

For the LMM analyses, we report coefficient estimates (*b*), standard errors (*SE*), and *t* values for each predictor. We assume that an absolute *t* value of 1.96 or greater indicates significance at the 0.05 alpha level (Baayen et al., 2008). For the BMM analyses, we report coefficient estimates of effect size (*b*), estimate errors (*SE*), and the 95% credible interval (CrI; i.e., under the model assumptions, there is a 95% probability that the parameter estimate is contained in this interval) for each predictor. If zero lies outside the credible interval, then we conclude there is sufficient evidence to suggest the estimate is different from zero.

### 2.2.3. Results

#### *2.2.3.1. Analysis of Response Times*

On average, participants responded 136 ms (see Fig. 2) before the end of the speaker's utterance, and 92% of the responses occurred within 1000 ms of the speaker's turn-end (see Fig. 3).

Figure 2. Observed means of response times (left) and precision (right) for the four conditions in Experiment 1. Error bars represent ± 1 standard error from the mean.

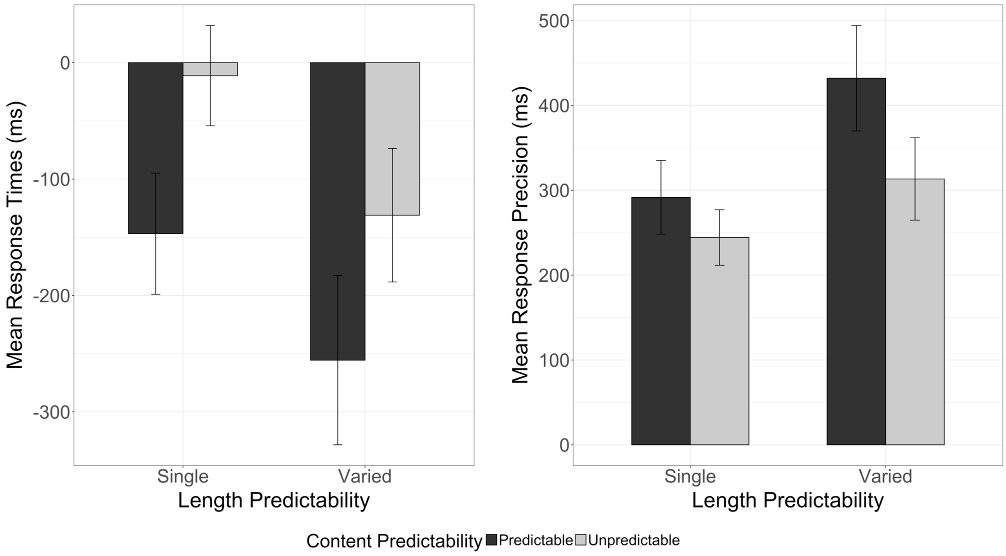
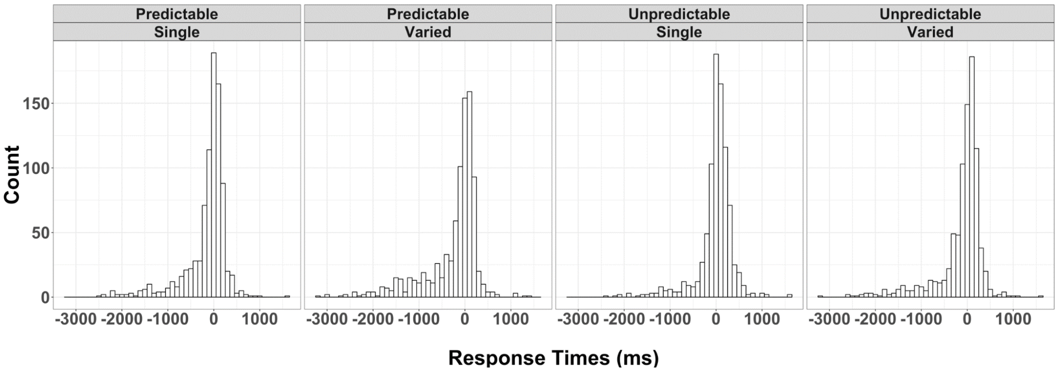


Figure 3. The distribution of observed response times in the four conditions in Experiment 1. Trials are placed into 100 ms time bins.



We found no significant effects of Content predictability (*b* = -28.31, *SE* = 29.10, *t* = -0.97) or Length predictability (*b* = -19.25, *SE* = 34.00, *t* = -0.55), and no interaction between the two (*b* = -8.57, *SE* = 50.15, *t* = 0.17; see the Appendix for full models). In contrast, Question Duration was a negative predictor of response times (*b* = -152.17, *SE* = 15.04, *t* = -10.12): Longer questions elicited earlier responses than shorter questions. Although there is a numerical difference in response times and response precision between the conditions in Fig. 2, note that these means are not adjusted for Question Duration, and our models show that this variable explains any differences in the observed means between conditions.

#### *2.2.3.2. Precision Analysis*

Participants responded on average 303 ms away from the end of the speaker's turn (see Fig. 2 for a breakdown by condition). We found no evidence that either Content predictability (*b* = -1.03, *SE* = 1.10, CrI[-0.22, 0.16]), Length predictability (*b* = -1.04, *SE* = 1.12, CrI[-0.25, 0.17]), or the interaction between the two (*b* = -1.28, *SE* = 1.20, CrI[-0.60, 0.10]) affected the scale parameter of the distribution. However, Question Duration had a positive effect on scale (*b* = 1.19, *SE* = 1.05, CrI[0.07, 0.27]), such that the spread of the distribution was greater when questions were longer.

### 2.2.4. Discussion

In Experiment 1, we investigated whether turn-end prediction plays a role in conversational turn-taking, as predicted by the late-planning hypothesis (e.g., Sjerps & Meyer, 2015; see Table 2). Specifically, we examined whether listeners predict the speaker's turn-end by predicting its content and length independently of one another (length-prediction hypothesis*;* Magyari & De Ruiter, 2012), or whether they predict length only if content is predictable (content-prediction hypothesis; Magyari et al., 2014). Recall that the early-planning hypothesis assumes that turn-end prediction does not play a role in turn-taking, and so makes no predictions for this task (see Table 2).

Inconsistent with the late-planning hypothesis, we found no effects of content or length predictability when analyzing the precision of participants' button-press responses. Instead, responses were influenced by question duration: Longer questions elicited less precise (and earlier) responses than shorter questions, as in previous research using the button-press paradigm (e.g., De Ruiter et al., 2006). There were also no content and length effects on signed response times; this contrasts with previous findings using the button-press paradigm (e.g., Magyari et al., 2014; Magyari & De Ruiter, 2012), which have shown that listeners respond earlier to predictable than unpredictable turns, even when conditions are matched for average duration.

This duration effect could be interpreted in line with previous research using reaction time experiments (see also Magyari, De Ruiter, & Levinson, 2017), which has found that response times are longer when the interval between a warning signal (alerting participants to the forthcoming reaction stimulus) and the reaction stimulus is shorter (e.g., Näätänen, 1971). When the utterance is longer, the interval between the warning signal and the reaction stimulus (i.e., between turn onset and turn-end) is also longer, and since the probability of the reaction stimulus (the turn-end) occurring continuously increases (Sanders, 1966), the listener is more likely to respond earlier when the utterance is longer in duration.

Another possibility is that longer turns elicit earlier responses because they typically contain more points of possible turn completion (see Sacks et al., 1974), and the listener may simply be more likely to mistake one of these points of completion for the actual turn-end. For example, consider the long question (2761 ms) *Did The Titanic sink after hitting an iceberg*?. It contains two plausible completion points: One after *sink*, and another after *iceberg*. Now compare it to the short question (1729 ms) *Are dogs your favorite animal*?, which contains only one plausible completion point (after *animal*) that coincides with the end of the question. Listeners may respond earlier to the first turn because there is an additional point of possible turn completion, before the actual turn-end.

In sum, the results of Experiment 1 did not provide any evidence to suggest that participants used either content or length predictability to determine the speaker's turn-end. Following Dienes (2014), we compared the null effect of content predictability with a hypothesized effect size distribution ranging between 0 and twice the mean condition difference reported by Magyari et al. (2014): 209 ms. The resulting Bayes factor was less than 0.33 (B = 0.11), indicating strong evidence in favor of the null hypothesis. (Note that we could not compute Bayes factors for the effect of Length predictability because we lack a measure of effect size.) These findings are more consistent with the early-planning hypothesis, which suggests listeners use predictions of turn content to prepare a response, but not to predict the speaker's turn-end. Since our conclusions are based on null results, however, we conducted Experiment 2 (a question-answering task) to test further predictions of the latter hypothesis, namely that listeners use content predictions to prepare a response as early as possible.

## 2.3. Experiment 2

Experiment 2 was identical to Experiment 1, with the exception that participants verbally answered each question either *yes* or *no*. If the early-planning hypothesis is correct, then we expected participants to answer earlier when question content was predictable rather than unpredictable. Since we found no evidence to suggest listeners used content or length predictability to predict turn-endings in Experiment 1, we did not predict any effects of content or length predictability on the precision of participants' verbal responses.

### 2.3.1. Method

#### *2.3.1.1. Participants*

Thirty new participants from the same population as in Experiment 1 (4 males, *M*age = 19.43) participated on the same terms.

#### *2.3.1.2. Materials and Procedure*

The materials and procedure were identical to those used in Experiment 1, with the exception that participants were told: "Answer as quickly as possible. Do not wait until the speaker has finished the question and has stopped speaking. Instead, you should answer as soon as you expect the speaker to finish the question". Thus, participants were encouraged to prepare a response as soon as possible (rather than simply wait for the speaker to finish) and articulate it close to the speaker's turn-end. Participants spoke into the microphone, and playback stopped as soon as a response was recorded using a voicekey.

### 2.3.2. Data Analysis

Response times and precision were calculated using the same procedure as Experiment 1. Of the 3468 responses, 188 (5.42%) were discarded because they could not be categorized as *yes* or *no*. We removed a further 12 (0.35%) response times greater than 10000 ms, as they were clear outliers. We then replaced 45 response times (1.37%) at the upper limit and 27 (0.37%) at the lower limit.

We fitted models using the same procedure as in Experiment 1. However, we included two further predictors to account for possible answer characteristics. *Yes* responses are usually produced faster than *no* responses (e.g., Strömbergsson, Hjalmarsson, Edlund, & House, 2013), and so we included Answer Type (reference level: no vs. yes) in our analyses. Since some of our questions were fact-based (e.g., *Did The Titanic sink after hitting an iceberg*?) while others were opinion-based (e.g., *Are dogs your favorite animal*?) we also included Agreement, which was the absolute difference between the percentage of participants who answered yes and the percentage who answered no. We assume that fact-based questions are likely to have a clear answer, and so Agreement will be high (a maximum of 100 when all participants provide the same answer). Thus, participants may need less time to determine what to say. For opinion-based questions, however, both *yes* and *no* are equally plausible answers, and thus Agreement will be low (a minimum of 0 when half of the participants answer *yes* and half answer *no*). As a result, participants may need more time to decide what to say.

### 2.3.3. Results

#### *2.3.3.1. Response Time Analysis*

On average, participants responded 379 ms after the end of the speaker's turn (see Fig. 4), and 90% of responses occurred within 1000 ms of the speaker's turn-end (see Fig. 5).

Figure 4. Observed means of response times (left) and precision (right) for the four conditions in Experiment 2. Error bars represent ± 1 standard error from the mean.

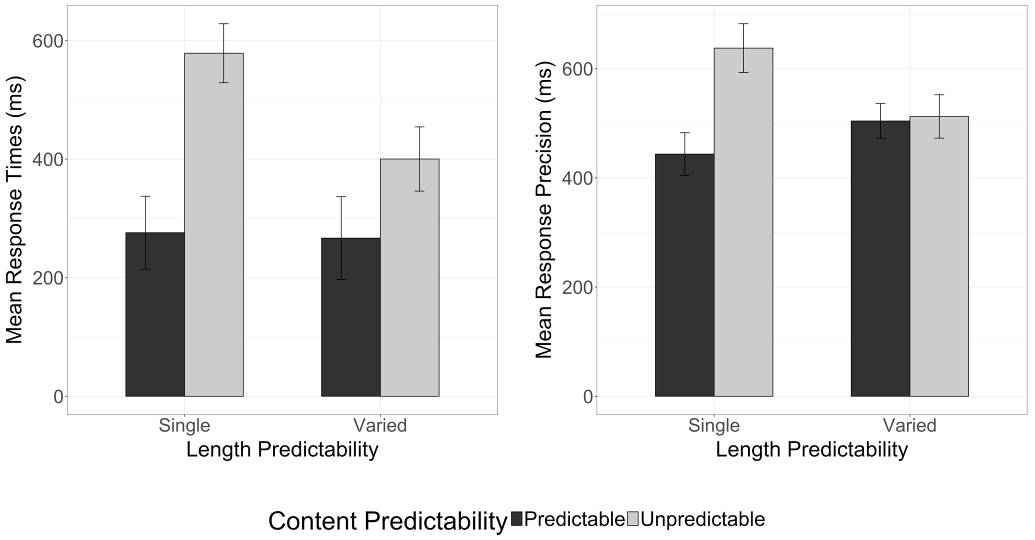
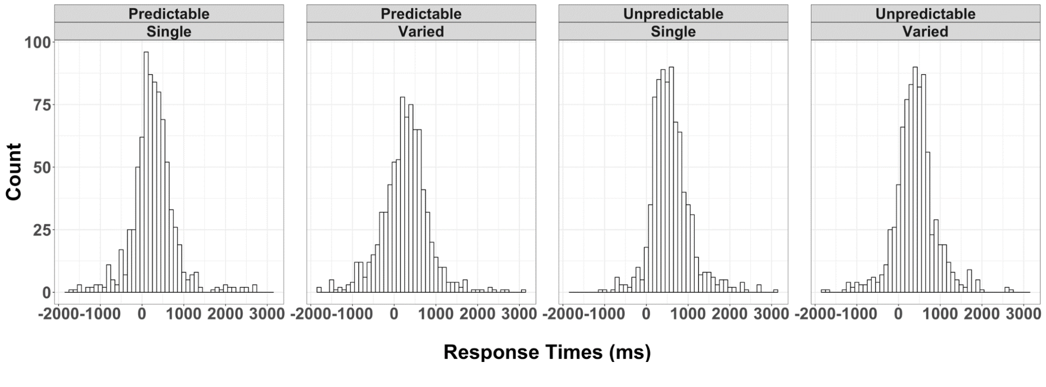


Figure 5. The distribution of observed response times in the four conditions in Experiment 2. Trials are placed into 100 ms time bins.



Participants answered earlier when content was predictable rather than unpredictable (*b* = -153.01, *SE* = 34.08, *t* = -4.49). However, there was no effect of Length predictability (*b* = 10.89, *SE* = 33.25, *t* = 0.33), and no interaction between Content and Length predictability (*b* = -110.21, *SE* = 63.75, *t* = -1.73). Inconsistent with previous research (e.g., Strömbergsson et al., 2013), response times were not affected by Answer Type (*b* = -21.86, *SE* = 16.46, *t* = -1.33): Participants were equally fast to respond *yes* and *no,* which may suggest that having participants interact with a pre-recorded speaker, rather than an actual interlocutor, reduces the social bias against "no" responses. However, Agreement was a significant negative predictor of response times (*b* = -55.21, *SE* = 15.17, *t* = -3.64): As expected, questions with higher agreement elicited earlier response times than those with lower agreement. In addition, longer questions elicited earlier responses than shorter questions (*b* = -72.88, *SE* = 17.25, *t* = -4.23), as in Experiment 1.

### *2.3.3.2. Precision Analysis*

On average, participants answered 509 ms away from the end of the speaker's turn (see Fig. 4 for a breakdown by condition). We found no evidence for an effect of either Content predictability (*b* = 1.05, *SE* = 1.13, CrI[-0.17, 0.28], Length predictability (*b* = 1.02, *SE* = 1.08, CrI[-0.14, 0.18], or their interaction (*b* = -1.20, *SE* = 1.15, CrI[-0.47, 0.09]. Precision was not influenced by Answer Type (*b* = -1.01, *SE* = 1.04, CrI[-0.10, 0.07] or Agreement (*b* = -1.06, *SE* = 1.03, CrI[-0.13, 0.00], but the spread of the distribution was greater when questions were longer in duration (*b* = 1.16, *SE* = 1.04, CrI[0.08 0.22]), as in Experiment 1.

### 2.3.4. Comparison analysis with Experiment 1

To determine whether the effect of content predictability in Experiment 2 was significantly different from Experiment 1, we conducted a cross-experiment comparison. We used the same analysis structure as in Experiment 2, but included an interaction between Content predictability, Length predictability, and Experiment (reference level: question-answering vs. button-pressing). Experiment was contrast coded (-0.5, 0.5), centered, and included as by-items random slopes. Since the size of the estimates suggested that Question Duration had a larger effect in Experiment 1 (*b* = -152.17) than 2 (*b* = -72.88), we included a Question Duration by Experiment interaction in the fixed effects structure of the model. Although we did not include Answer Type (yes or no) as a main effect because this variable was participant-specific (i.e., different participants answered yes or no to different items), we did include Agreement, since this variable was item-specific.

Importantly, when analyzing response times, we found a significant effect of Content predictability (*b* = -86.88, *SE* = 29.75, *t* = -2.92), Experiment (*b* = -491.56, *SE* = 79.38, *t* = -6.19), and a significant interaction between the two (*b* = 156.80, *SE* = 39.69, *t* = 3.95), confirming that Content predictability affected the timing of participants' verbal responses more than the timing of their turn-end predictions. In addition, there was no effect of Length predictability, and this predictor did not interact (either two-way or three-way) with any other predictors (all *ts* $\leq$1.96).

When analyzing the precision of participants' responses, we found an effect of Experiment (*b* = -1.90, *SE* = 1.22, CrI[-1.04, -0.24], but no effect of Content predictability (*b* = -1.05, *SE* = 1.07, CrI[-0.19, 0.10]), Length predictability (*b* =

-1.01, *SE* = 1.07, CrI[-0.14, 0.12]), and no interaction between any of these predictors (all CrIs included 0). Response times and precision were influenced by Agreement and Question Duration in the same way as in the individual analyses; in addition, Agreement had a negative influence on the precision of responses in the comparison analysis (*b* = -1.08, *SE* = 1.03, CrI[-0.13, -0.03]), even though it did not in the individual experiment analyses. These results suggest that the lack of predictability effects on the precision of participants' responses was comparable in the question-answering and button-pressing tasks. Along with the individual experiment analyses, these results confirm there was an effect of content predictability in the question-answering task, but not in the button-pressing task. Thus, participants used content predictions to prepare their response, but not to predict the speaker's turn-end.

### 2.3.5. Discussion

In Experiment 2, we investigated whether early response preparation occurs during turn-taking. Participants answered earlier when question content was predictable rather than unpredictable, suggesting they used predictions of turn content to prepare a verbal response. In contrast, we found no effects of content or length predictability on the precision of participants' responses. Together with Experiment 1 and our cross-experiment comparisons, these results suggest that listeners in our experiments used content predictions to prepare their verbal response as early as possible but not to predict the turn-end, and are thus consistent with the early-planning hypothesis.

However, in both Experiments 1 and 2, our measures of content predictability were not comparable across the single and varied length conditions. Since we used multi-word completions in the varied conditions, the predictability of completions was assessed at an earlier point in the varied than in the single conditions. For example, the unpredictable varied question *Do most students finish their exams on time*? was cut off three words before question end (*Do most students finish their…*) in the pre-test, whereas the unpredictable single question *Do you enjoy going to the supermarket?* was cut off just one word before question end (*Do you enjoy going to the…*). But the content predictability of the utterance may well increase with each additional word the speaker produces. For instance, the listener cannot predict what the speaker will say after the words *Do most students finish their…* (and so the predictability of question content is fairly low at this point), but may be able to predict *time* after hearing *Do most students finish their exams on…*).

Indeed, when we conducted a cloze post-test to assess the content predictability of the final word of the questions in the varied conditions, in which 33 participants from the same population as Experiment 1 (8 males; *M*age = 20.15) completed the same procedure as previous pre-tests, we found that stimuli in the two varied conditions had significantly higher content predictability (predictable varied completion cloze: 76%, unpredictable varied completion cloze: 68%; predictable varied completion LSA: 0.83, unpredictable varied completion LSA: 0.73) than those in the unpredictable single condition (completion cloze: 4%; completion LSA: 0.16; all *p*s < .001). Thus, even though the predictable and unpredictable single conditions demonstrate that listeners can use content predictions to prepare their responses early, our measures of content predictability in the varied conditions were not comparable to those in the single conditions. This may have affected our length predictability manipulation, and so we conducted two further experiments (Experiments 3 and 4) in which all stimuli had single word completions to provide a further test of the length prediction hypothesis.

## 2.4. Experiment 3

Experiment 3 was identical to Experiment 1, in that participants were instructed to press a button when they thought the speaker had reached the end of their turn, but we selected single word completions for all stimuli to ensure content predictability was comparable across the conditions. We also discarded the predictable varied condition from Experiment 1 because most of these stimuli were completed with a single word most of the time, and so a single word completion would have been predictable in this condition.

Importantly, discarding the predictable varied condition does not affect our ability to disentangle late from early-planning, as we can still examine effects of content predictability across the button-press and the question-answering paradigm. It also does not affect our ability to determine whether participants predicted the speaker's turn-end by predicting the length of the speaker's utterance separately from its content, as we can still compare the two unpredictable content conditions. The content-prediction hypothesis predicts no difference in response precision in the two unpredictable content conditions; the length-prediction hypothesis predicts that responses should be more precise for unpredictable utterances whose length is predictable (i.e., unpredictable single condition) rather than unpredictable (i.e., unpredictable varied condition).

To minimize any confounding effect of Question Duration (as occurred in Experiment 1), we followed Magyari et al. (2014) and matched the average duration of the three stimulus conditions. Since we also used the same stimuli in Experiment 4, we matched the average Agreement of the three conditions.

### 2.4.1. Method

#### *2.4.1.1. Participants*

Thirty new native English speakers (10 males; *M*age = 22.20) at the University of Edinburgh participated on the same terms as previous experiments.

#### *2.4.1.2. Materials*

We constructed 141 question fragments, sometimes by re-using materials from Experiment 1. Note that we pre-tested both old and new fragments to ensure consistency across the item set. We selected completions for these fragments using the same pre-test procedure as in Experiment 1, with 33 new native English speakers (2 males, *M*age = 20.03 years). Using these responses, we selected 28 stimuli for each of the three conditions (84 stimuli in total).

We calculated content and length predictability as in Experiment 1. However, we selected single word completions for all fragments in all conditions. This completion length was used by at least 90% of participants in the single conditions, and by no more than 72% of participants in the unpredictable varied condition (see Table 4). Questions in the predictable and unpredictable single conditions were matched for average completion length variance ($p$ = .15), and both conditions had lower variance than questions in the unpredictable varied condition (all *p*s < .001; see Table 4).

Table 4. The means (and standard deviations) of our measures of content predictability, length predictability, difficulty, plausibility, answer agreement, and duration (ms) for stimuli in Experiments 3 and 4. The final column provides the number of utterances characterized by a pitch downstep in each condition.

| | Predictable Single | Unpredictable Single | Unpredictable Varied |
|---|---|---|---|
| Average Completion Length Variance | 0.03 (0.04) | 0.05 (0.05) | 0.88 (0.59) |
| Completion Length Cloze[a] | 98% (3%) | 97% (3%) | 38% (21%) |
| Question Fragment LSA[b] | .90 (.11) | .37 (.12) | .34 (.10) |
| Completion LSA[c] | .94 (.07) | .15 (.07) | .20 (.14) |
| Completion Content Cloze[d] | 91% (9%) | 5% (2%) | - |
| Question Fragment Entropy[e] | 0.43 (0.37) | 2.96 (0.68) | - |
| Question Difficulty[f] | 6.34 (0.52) | 6.00 (0.76) | 6.21 (0.47) |
| Question Plausibility[g] | 5.78 (0.64) | 5.58 (0.56) | 5.68 (0.52) |
| Answer Agreement | 53% (36%) | 37% (27%) | 43% (27%) |
| Question Duration (ms) | 2284 (632) | 2021 (560) | 2031 (489) |
| Downstepped utterances | 23/28 | 22/28 | 15/28 |

[a] Percentage of participants who provided the word length of the selected completion used in the main experiment (a single word in the single conditions; multiple words in the varied conditions) as a continuation in the cloze task.

[b] Average over all completion comparisons for that particular fragment.

[c] Average over comparisons between the selected completion and all other completion.

[d] Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

[e] Entropy of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.

[f] Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, while 7 indicated that the question was very plausible/easy to answer.

Stimuli in the predictable single condition had higher fragment LSA than the two unpredictable content conditions (all *p*s < .001). In addition, the predictable single condition had higher cloze probability and lower entropy than the unpredictable single condition (all *p*s < .001). The LSA values for the two unpredictable conditions were matched (all *ps* > .13; see Table 4).

We matched the mean difficulty, plausibility, and answer agreement (all *p*s > .09) of the three conditions using data from a separate pre-test, in which participants (31 native English speakers; 5 males, *M*age = 20.58) answered each question either *yes* or *no* and rated the difficulty and plausibility of questions, as in Experiment 1. Questions were recorded by the same native English speaker as in Experiment 1, and were matched for average duration (all *p*s > .21; see Table 4). When analyzing the pitch contours of these questions, six (7%) had creaky voice, all had falling boundary tones, and sixty (71%) had a downstep in pitch (see Table 4). Both judgments were again validated the same second coder as in Experiment 1, who rated 25% of the stimuli. This resulted in a Cohen's kappa of 1 for boundary tone judgements and .72 for downstep judgements, which is considered "good" agreement (see Cicchetti, 1994; Landis & Koch, 1977). Note that, if listeners use downsteps to determine the speaker's turn-end (e.g., Cutler & Pearson, 1986), then we would expect them to be more precise at timing their response in the unpredictable varied condition (where there are more downsteps) than in either the unpredictable or the predictable content single conditions. However, this is the opposite of the predictions made by the content- or length-prediction hypotheses.

#### *2.4.1.3. Procedure*

The procedure was identical to Experiment 1, except that breaks occurred after every 28 stimuli.

### **2.4.2. Data Analysis**

Response times and precision were analyzed as in Experiment 1. We replaced 12 response times (0.48%) above the upper limit, and 66 (2.62%) below the lower limit with the cut-off value. Data analysis, predictors, and random effects structure were identical to those used in Experiment 1. However, we defined two orthogonal Helmert contrasts to capture effects of Content and Length predictability. The Content contrast compared the mean of the two unpredictable conditions (1/3) to the predictable condition (-2/3, reference level), and the Length contrast compared the unpredictable varied condition (0.5) to the unpredictable single condition (-0.5, reference level). Since the two contrasts are orthogonal, no interaction term was included. Even though we balanced Question Duration, we still included it as an additional main effect to ensure our results could not be attributed to any residual differences. All predictors were centered.

### 2.4.3. Results and Discussion

#### *2.4.3.1. Analysis of Response Times*

Participants responded 117 ms before the end of the speaker's turn (see Fig. 6) and 93% of responses occurred within 1000 ms of the end of the speaker's question (see Fig. 7).

Figure 6. Observed means of response times (left) and precision (right) for the three conditions in Experiment 3. Error bars represent ± 1 standard error from the mean.

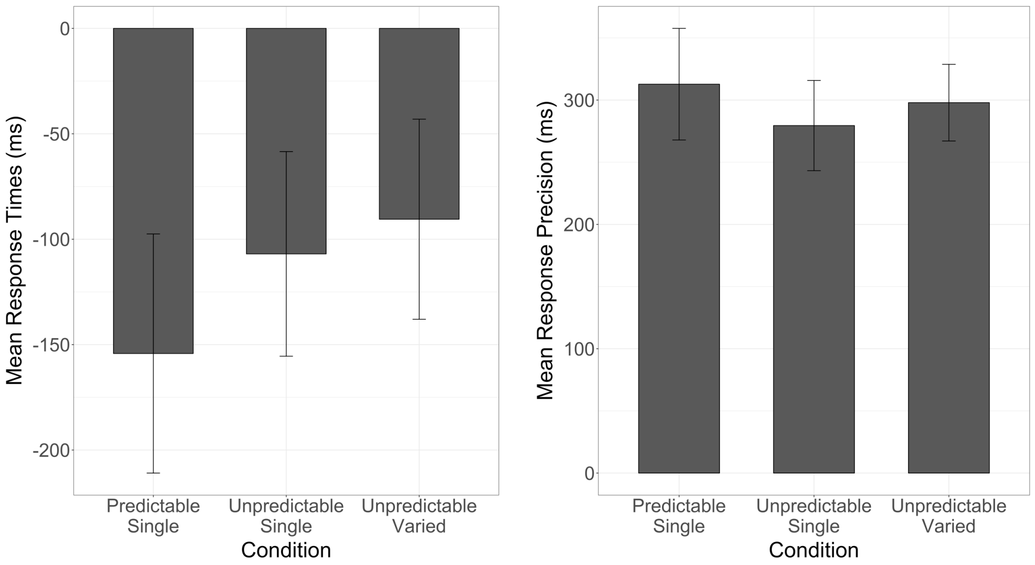
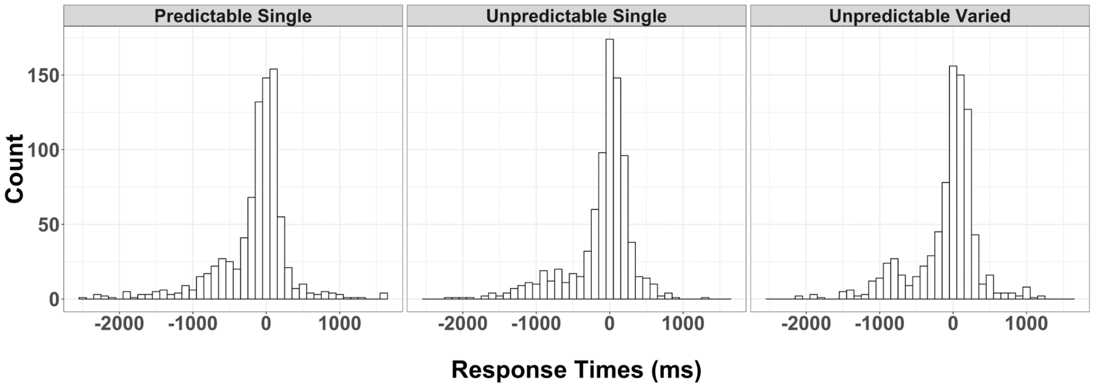


Figure 7. The distribution of observed response times in the three conditions in Experiment 3. Trials are placed into 100 ms time bins.



As in Experiment 1, we found no significant effect of Content (*b* = 0.39, *SE* = 35.60, *t* = 0.01) or Length predictability (*b* = 18.75, *SE* = 41.74, *t* = 0.45). The Bayes factor for the null effect of content predictability was 0.05, again indicating strong evidence in favor of the null hypothesis. Question Duration was still a negative predictor of response times (*b* = -125.00, *SE* = 41.74, *t* = -8.98).

### *2.4.3.2. Precision Analysis*

Participants responded 297 ms away from the end of the speaker's question on average (see Fig. 6). We found no evidence for an effect of either Content predictability (*b* = 1.26, *SE* = 1.16, CrI[-0.07, 0.53]) or Length predictability (*b* = 1.11, *SE* = 1.30, CrI[-0.41, 0.62]). However, the spread of the distribution was again greater when questions were longer in duration (*b* = 1.26, *SE* = 1.05, CrI[0.13, 0.32]). These results are consistent with Experiment 1, and provide no support for the idea that listeners used content or length predictability to predict the speaker's turn-end.

## 2.5. Experiment 4

Experiment 4 was identical to Experiment 2, in that participants verbally answered each question either *yes* or *no*, but we used the same stimuli from Experiment 3. If participants use content predictions to prepare a verbal response, then we expect them to answer earlier when question content is predictable rather than unpredictable. Since we found no evidence to suggest listeners used content or length predictability to determine the end of the speaker's turn in any of the previous experiments, we did not expect either of these variables to influence response precision.

### 2.5.1. Method

#### *2.5.1.1. Participants*

Thirty new participants from the same population in the previous three experiments (10 males; *M*age = 22.20) took part on the same terms.

#### *2.5.1.2. Materials and Procedure*

The materials were identical to those used in Experiment 3, and the procedure was identical to that used in Experiment 2.

### 2.5.2. Data Analysis

We discarded 39 responses (1.58%) because they could not be clearly categorized as *yes* or *no*. We discarded nine (0.36%) response times greater than 10000 ms, and then replaced 39 response times (1.58%) at the upper limit and 30 (1.21%) at the lower limit. We analyzed response times and precision using the same procedure as Experiment 3, but in addition we also included Answer Type (reference level: no vs. yes) and Answer Agreement as main effects.

## 2.5.3. Results and Discussion

### *2.5.3.1. Analysis of Response Times*

Participants responded 484 ms after the end of the speaker's turn (see Fig. 8) and 89% of responses occurred within 1000 ms of question end (see Fig. 9).

Figure 8. Observed means of response times (left) and precision (right) for the three conditions in Experiment 4. Error bars represent ± 1 standard error from the mean.

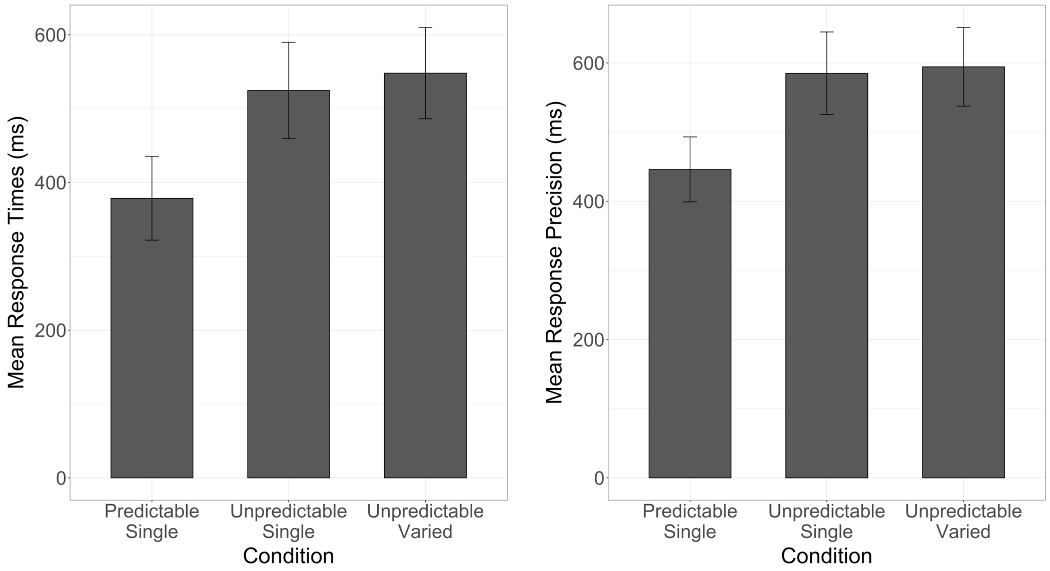
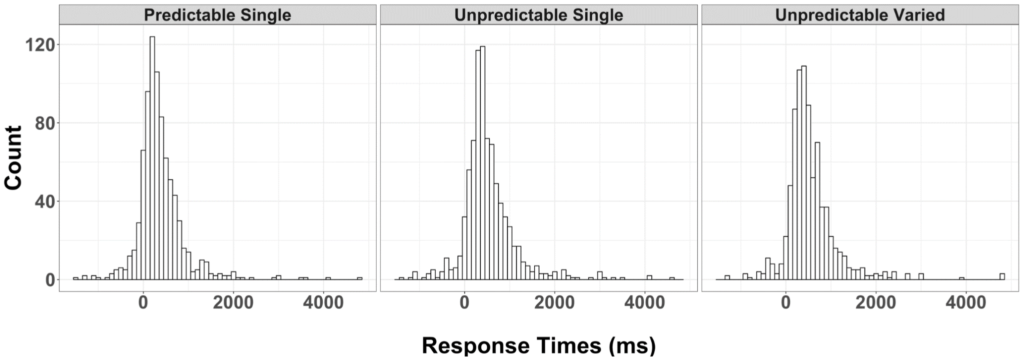


Figure 9. The distribution of observed response times in the three conditions in Experiment 4. Trials are placed into 100 ms time bins.



Participants answered earlier when question content was predictable rather than unpredictable (*b* = 95.78, *SE* = 34.54, *t* = 2.77). However, there was no effect of Length predictability (*b* = 28.19, *SE* = 36.81, *t* = 0.77). These results replicate Experiment 2, and suggest that participants prepared their answer as early as possible.

Unlike Experiment 2, participants answered *yes* earlier than *no* (*b* = -143.43, *SE* = 19.92, *t* = -7.20). This replicates previous studies (e.g., Stivers et al., 2009; Strömbergsson, et al., 2013) and suggests that the lack of an effect of Answer Type in Experiment 2 cannot be attributed to the fact that our participants interacted with a pre-recorded speaker rather than an actual interlocutor. In addition, participants answered questions with higher agreement earlier than those with lower agreement (*b* = -35.81, *SE* = 15.66, *t* = 2.29). Finally, questions longer in duration elicited earlier response times than those shorter in duration (*b* = -59.85, *SE* = 15.67, *t* = -3.82). Together with Experiment 2, these results suggest that Answer Type, Agreement, and Question Duration all influence response times during a question-answering paradigm.

#### *2.5.3.2. Precision Analysis*

Participants responded 542 ms away from the end of the speaker’s question (see Fig. 8). Response precision was not influenced by Content predictability (*b* = 1.13, *SE* = 1.11, CrI[-0.08, 0.33]), Length predictability (*b* = 1.01, *SE* = 1.16, CrI[-0.28, 0.31]), Answer Type (*b* = 1.00, *SE* = 1.05, CrI[-0.09, 0.09]), or Answer Agreement (*b* = 1.02, *SE* = 1.04, CrI[-0.05, 0.09]). However, the spread of the distribution was greater when questions were longer in duration (*b* = 1.12, *SE* = 1.04, CrI[0.04, 0.18]). These results replicate Experiment 2, and suggest participants did not time response articulation by predicting the content or the length of the speaker’s question.

### 2.5.4. Comparison analysis with Experiment 3

As in Experiments 1 and 2, we conducted a cross-experiment comparison between Experiments 3 and 4. We used the same analysis structure as in the previous cross-experiment comparisons, but with predictors defined as in Experiment 4. Recall that Content and Length predictability were implemented as orthogonal contrasts in Experiment 4; therefore, we included two three-way interactions between Content predictability, Experiment, and Question Duration and between Length predictability, Experiment, and Question Duration, but no four-way interaction.

We could not analyze the precision of participants’ responses because the model did not converge ( $\widehat{R}$ values > 1.1), but note that we found no effects of either Content or Length predictability on precision in either Experiment 3 or 4. Below, we report only a cross-experiment comparison of the analysis of response times.

Importantly, when analyzing response times, we found no significant effect of Content predictability (*b* = 40.15, *SE* = 33.24, *t* = 1.21) or Length predictability (*b* =52.22, *SE* = 61.83, *t* = 0.84). There was a significant effect of Experiment (*b* = -597.21, *SE* = 15.33, *t* = -38.97), such that participants responded earlier in the button-press than question-answering task. As in Experiment 1, there was an interaction between Content predictability and Experiment (*b* = -132.56, *SE* = 36.08, *t* = -3.67). But there was no interaction between Length predictability and Experiment (*b* = -13.52, *SE* = 67.79, *t* = -0.20). Response times were influenced by Answer Agreement in the same way as in the individual experiment analyses (*b* = -36.80, *SE* = 14.22, *t* = -2.59). Together with the individual analyses, these results suggest that the effect of content predictability was stronger in the question-answering than button-pressing experiment. In other words, these results provide further evidence to suggest listeners used content predictions to prepare a verbal response, but not to predict the speaker's turn-end.

## 2.6. General Discussion

In four experiments, we used button-press (Experiments 1 and 3) and question-answering (Experiments 2 and 4) tasks to investigate how interlocutors use prediction to achieve finely coordinated turn-taking. We contrasted two different hypotheses: (i) the early-planning hypothesis, which proposes that listeners use content predictions to prepare an early response but not to predict the speaker's turn-end (e.g., Levinson & Torreira, 2015), and (ii) the late-planning hypothesis, which proposes that listeners use content predictions (content-prediction hypothesis) and possibly length predictions (in number of words; length-prediction hypothesis) to determine the speaker's turn-end, and only begin preparation close to this moment (e.g., Sjerps & Meyer, 2015). In all experiments, we manipulated both the content (i.e., the predictability of the words of the speaker's turn) and length predictability (i.e., the predictability of the number of words needed to complete the turn) of simple *yes/no* questions.

There were no predictability effects on the precision of participants' button-presses or verbal responses (i.e., how closely participants responded to the speaker's turn-end), suggesting that listeners did not use linguistic information (either about content or length) to predict the speaker's turn-end. However, we did find effects of content predictability on response times in the question-answering tasks: Participants answered earlier when the final word(s) of the question were predictable (e.g., *Are dogs your favorite animal*?) rather than unpredictable (e.g., *Do you enjoy going to the supermarket*?). These results are consistent with findings from studies in language comprehension, which have shown that listeners can use the content of the speaker's utterance to predict how it continues (e.g., Altmann & Kamide, 1999), and suggest that listeners used such predictions to prepare their own response early during the speaker's turn.

Our findings are consistent with previous research that supports early-planning during turn-taking (e.g., Barthel et al., 2016, 2017; Bögels et al., 2015) and suggest that listeners used content predictions to prepare their response early, but not to predict when they could launch articulation of this response. In contrast, our findings are inconsistent with the late-planning hypothesis, which suggests that listeners delay preparation until they know that they will soon have the opportunity to launch articulation. Specifically, Sjerps and Meyer (2015) found that listeners delayed preparation until near the end of the speaker's utterance. However, it may be that this discrepancy is due to their use of the dual-task paradigm: If participants had prepared a response early then they would have had to carry out three simultaneous tasks (i.e., comprehending the speaker's turn, preparing their own response, and finger tapping). Thus, their participants may have delayed preparation because they used cognitive resources to carry out an additional attention-demanding task, which is normally absent during conversation. Sjerps and Meyer addressed this issue in their second experiment, in which they found that participants looked towards to-be-named objects only shortly before producing their response. However, listeners may have given preference to looking for comprehension, and thus did not look earlier at the objects that they themselves had to name.

Our results are inconsistent with both the length-prediction hypothesis, which proposes that listeners predict the speaker's turn-end by predicting the length (in number of words) of the speaker's utterance, even when content is unpredictable (e.g., Magyari & De Ruiter, 2012), and the content-prediction hypothesis (Magyari et al., 2014), which instead suggests that listeners predict length only when content is predictable. However, there are a number of notable differences between our experiments and previous studies that have manipulated the content or length predictability of turns. First, neither Magyari and De Ruiter nor Magyari et al. included utterance duration as a control variable in their analyses. Duration was a strong predictor of response times in both of our button-press experiments (and those reported by De Ruiter et al., 2006): We found that questions longer in duration elicited less precise and earlier responses than those shorter in duration. Thus, it is possible that previous findings can be attributed to residual differences in duration, even if those studies matched the average duration of turns across conditions.

But other studies, which have fully controlled for duration, demonstrated turn-end prediction does play a role in turn-taking, and specifically that being able to understand the content of the speaker's utterance is important for determining the speaker's turn-end (e.g., De Ruiter et al., 2006; Riest et al., 2015). It is less clear, however, whether these studies demonstrate that the predictability of this content is important. In fact, Riest et al. found no difference between a condition in which participants could preview a transcript of the turn and one in which they were exposed to the turn for the first time. They interpreted this as evidence that speakers predicted the turn-end in both conditions, but it could also be interpreted as evidence that predictability does not affect how early participants respond in the button-press paradigm (there was no separate assessment of turn predictability, so it is difficult to determine how predictable the turns were when participants heard them for the first time).

Another difference between our study and previous ones is that our questions were produced by a pre-recorded speaker, while those in previous studies (e.g., De Ruiter et al., 2006) were taken from natural conversation. Thus, we may have failed to replicate their effects of content predictability because certain characteristics (i.e., changes in pitch, intonation, etc.) that are present in natural stimuli may have been absent in our recorded stimuli. We also note that both of our experiments used an explicit task, in which participants were encouraged to predict the speaker's turn-end (Experiments 1 and 3) and answer quickly (Experiments 2 and 4). But in natural conversation, listeners are unlikely to predict turn-ends explicitly or be aware of the explicit pressure to respond quickly. Nevertheless, these tasks allow us to tap into some of the mechanisms underlying coordination during turn-taking.

In sum, our results suggest that listeners can and do prepare their response early. Future research could explore what aspects of their response listeners prepare in advance. It is possible that they prepare the lexical content of their response and hold this response in an articulatory buffer until they can launch articulation (see Piai et al., 2015a). But assuming that production and comprehension share resources (e.g., Segaert et al., 2012), how does the listener manage to prepare and buffer a response while comprehending the speaker's unfolding turn? If the listener can predict what the speaker is going to say, then it may matter less that they fully comprehend the speaker's unfolding turn because they have already comprehended enough of the utterance to predict the speaker's message and prepare a response. Although some comprehension must be necessary, in case any prediction is inaccurate, the listener may manage the capacity demands of concurrent production and comprehension by allocating fewer resources to comprehending their interlocutor's turn. Further research could investigate this issue.

Regardless, listeners must still ensure they articulate their pre-prepared response at the appropriate moment. Listeners may rely on a number of mechanisms to do so (e.g., Bögels & Levinson, 2017; see also Wilson and Wilson, 2005). One possibility is that listeners launch articulation of their response reactively, after they have encountered one or more turn-final cues (e.g., falling boundary tone). This more reactive strategy (Duncan, 1972; Heldner & Edlund, 2010) may still be compatible with short inter-turn intervals because launching articulation does not take as long as preparing a response from scratch (articulation takes around 145 ms; Indefrey &

Levelt, 2004). Note that listeners are likely to be sensitive to a collection of such cues (e.g., Bögels & Torreira, 2015), and could use multiple cues to determine points of possible turn completion.

Importantly, these cues could work in parallel with a turn-end prediction mechanism, and this may well explain why turn-final cues are not necessarily perfect predictors of a speaker switch (e.g., Gravano & Hirschberg, 2011). For example, in instances when the listener is able to predict that the speaker will soon reach the end of their turn, they may allocate more processing resources to paying attention to possible turn-final cues, so that they are quicker to launch articulation when the speaker displays such cues. But in instances when such predictions are not possible, the listener may process such cues much less efficiently, resulting in longer gaps between turns.

In conclusion, we have shown that participants in a question-answering task were sensitive to the predictability of final words in questions: Participants answered earlier when such words were predictable rather than unpredictable. However, we found no evidence that participants used their ability to predict the final word to estimate when the speaker's turn would end. Thus, we conclude that content predictability helps listeners prepare a verbal response early, but does not help them determine when they should launch articulation of this response.

# 3. Study 2 Experiments 5 & 6: Using speech rate entrainment to time response articulation[5]

## 3.1. Introduction

Accurately predicting when future events will occur is important for many successful interactions. People often predict their partner's timing by tracking or entraining to temporal regularities in their partner's behavior, for example in their actions (e.g., Pecenka & Keller, 2011) or in their speech (e.g., Arnal & Giraud, 2012; Cummins, 2009). Accurate timing predictions are likely to be particularly important in natural conversation, in which speakers' contributions are so finely coordinated that there is little overlap or gap between their turns (around 200 ms on average; Stivers et al., 2009). But are the mechanisms underlying predictive timing in spoken dialogue similar to those used in language comprehension? In two experiments, we address this question and ask whether entrainment to speech rate during language comprehension influences the smooth timing of turns in language production.

There is much evidence that many of the representations used during language comprehension are the same as those used during language production. For example, comprehending word primes that are semantically or associatively related to the name of a target picture affects naming times during production (e.g., Alario, Segui, & Ferrand, 2000, Schriefers et al., 1990). Additionally, interlocutors often

[5] This chapter is based on a manuscript under review in Journal of Experimental Psychology: Learning, Memory, and Cognition (Corps, R. E., Gambi, C., & Pickering, M. J. (under review). How do listeners time response articulation during conversational turn-taking? The role of speech rate entrainment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.).

align their representations, such that they tend to repeat each other's choice of syntactic structure (Branigan et al., 2000) and referring expressions (e.g., Brennan & Clark, 1996). Finally, studies investigating syntactic repetition using fMRI (e.g., Menenti et al., 2011; Segaert et al., 2012) have shown that the same brain areas are affected in comprehension and production. Although it is unclear whether phonological representations are shared across modalities (see Gambi & Pickering, 2017), it certainly appears that representations of lexico-syntactic content (i.e., what the speaker is going to say) activated during comprehension can influence later production (and vice versa).

Recent findings suggest that listeners can use predictions of turn content to prepare their own response (e.g., Bögels et al., 2015). But in addition to deciding what they want to say, listeners in dialogue must also decide when they want to say it. Some studies suggest that the timing of events during comprehension can influence the timing of events during production (and the same is perhaps true across action and perception more generally; e.g., Knoblich, Butterfill, & Sebanz, 2011). For example, Jungers and Hupp (2009; Jungers et al., 2002) found that listeners were more likely to produce picture descriptions at a fast rate after hearing a prime sentence produced at a fast rate. Similar results were reported in dialogue by Schultz et al. (2016; see also Finlayson et al., 2012), who found that interlocutors' beat rates became mutually entrained during scripted turn-taking conversations: Participants produced their turn at a faster beat rate after their interlocutor produced their own turn at the same fast beat rate.

Together, these studies suggest that listeners entrain to their interlocutor's speech rate during comprehension, which can in turn influence the rate of their subsequent production. However, these studies have not tested whether such rate entrainment influences the timing with which listeners launch articulation of their turns. Speakers often vary in their speaking rates (e.g., Tauroza & Allison, 1990; Miller & Dexter, 1988; Miller, Grosjean, & Lomanto, 1984), and so listeners must take this information into account if they wish to produce their own turn at the appropriate moment (i.e., so they do not overlap or leave long gaps between utterances). If entrainment during comprehension can prime the timing of articulation, then we would expect listeners to produce their turn earlier when the previous speaker has produced their utterance at a faster rather than a slower rate.

Indeed, two theoretical accounts of conversational turn-taking argue that entrainment plays a key role in coordinating turns. First, Wilson and Wilson (2005) claimed that listeners entrain to (or track) an interlocutor's speech rate using cyclic neural oscillators, which are pools of neurons that synchronize to an external rhythm (Large & Jones, 1999). Indeed, much evidence suggests that neural oscillators underlie speech rate entrainment. For example, Zion Golumbic et al. (2013; see also Ding et al., 2017) recorded electrocorticographic (ECoG) activity in the auditory cortex while listeners attended to one of two speakers. They found that oscillations in both the high (75-150 Hz; associated with phrasal processing, see Giraud & Poeppel, 2012) and low (1-7 Hz; associated with phonemic and syllabic processing) frequency ranges tracked the signal of the attended speech. Further studies suggest that these oscillators are sensitive to the speaker's rate of syllable production. For example, Doelling et al. (2014; see also Ghitza, 2012) found that the correspondence between oscillatory activity and the speech signal was reduced when temporal fluctuations associated with syllable rate were removed. Entrainment was regained when these fluctuations were artificially reinstated by inserting silent gaps, so that the syllable rate of the manipulated turn was comparable to that of the natural turn.

Wilson and Wilson (2005) argued that conversational overlap is rare because interlocutors' oscillatory cycles are entrained in anti-phase, so that the listener's (i.e., the next speaker) readiness to produce a syllable is at a maximum when the current speaker's readiness is at a minimum (and vice versa). In the context of turn-taking, anti-phase means that listeners will be maximally ready to produce their turn half a syllable before or after the end of the speaker's turn. If listeners do not produce a response half a syllable before or after the end of a turn, then they will not be able to begin speaking again until after they have completed another oscillatory cycle (i.e., the duration of another syllable). This account therefore predicts that listeners will be maximally ready to produce their response half a syllable before or after the end of the speaker's turn, meaning that inter-turn intervals should be bimodally distributed around zero.

Although support for Wilson and Wilson's (2005) account can be drawn from studies demonstrating convergence of speech rate (e.g., Jungers & Hupp, 2009) and the duration of turn transitions during dialogue (e.g., Street, 1984), others have found that speech rate convergence does not influence the duration of inter-turn intervals (see Finlayson et al., 2012). Furthermore, there is no evidence to support Wilson and Wilson's argument that interlocutors' oscillatory cycles are in anti-phase. For example, Beňuš (2009) tested the oscillator theory using 12 dyadic conversations between people playing computer games from the Columbia Games Corpus. Turn intervals were unimodally (rather than bimodally) distributed, with a peak around

100-200 ms, which is consistent with research demonstrating that turn-intervals typically average 0-200 ms (Stivers et al., 2009).

In a second theoretical account, Garrod and Pickering (2015) also proposed that speech rate entrainment affects the duration of inter-turn intervals. Specifically, they argued that listeners use entrainment to predict the rate of the speaker's forthcoming syllables. This prediction then affects when listeners launch articulation, such that turn transitions should be shorter when the speaker produces their turn at a faster than a slower rate, because listeners should predict that they can launch articulation earlier. Unlike Wilson and Wilson (2005), however, this account does not make any claim about interlocutors' cycles being in anti-phase, and instead allows for many other factors to affect the duration of turn-intervals. In fact, research suggests that determining that the speaker is about to stop speaking likely depends on a number of mechanisms, such as predicting the speaker's turn-end (e.g., De Ruiter et al., 2006) and reacting to turn-final cues (such as downstepping, lengthening of the final word, or a drop in pitch; e.g., Bögels & Torreira, 2015; Gravano & Hirschberg, 2011), which suggest that the speaker is about to stop. Although turn-end prediction and turn-final cues are not the focus of our experiments, we controlled for their presence to ensure they did not affect the duration of inter-turn intervals.

Consistent with Garrod and Pickering's (2015) account, a number of studies suggest that once listeners have entrained to their interlocutor's speech rate, they predict that the speaker's forthcoming syllables will continue at the same rate. In one study, Dilley and Pitt (2010; see also Pitt, Szostak, & Dilley, 2016) either expanded (by a factor of 1.9) or compressed (by a factor of 0.6) the rate of the context surrounding a co-articulated single-syllable function word (e.g., *or* in *Deena doesn't* *have any leisure or time*). When context rate was slowed, listeners often failed to perceive this function word (*leisure or time* was perceived as *leisure time*); when context rate was speeded, listeners tended to erroneously perceive an absent function word (*leisure time* was perceived as *leisure or time*). Dilley et al. (2013) reported a similar pattern of results with reduced syllables, suggesting that this effect is not limited to function words. This disappearing-syllable effect is thought to occur because the listener has entrained to the speaker's syllable rate (but see Cummins, 2012) and predicts that future syllables will continue to be produced at the same rate. This prediction then causes the listener to adopt the interpretation that is more compatible with the predicted rate, leading to the loss or insertion of a syllable. In support of this interpretation, Kösem et al. (2017) found that low frequency activity in the auditory cortex entrained to the context rate of a sentence and was sustained after a rate change occurred.

Similar results have also been found over longer timescales. Using the same procedure as Dilley and Pitt (2010), Baese-Berk et al. (2014; see also Morrill, Dilley, McAuley, & Pitt, 2014) manipulated both the speech rate of individual utterances (the distal rate) and the average speech rate of utterances across the whole experiment (the global rate). They found that participants were less likely to perceive a function word when the context rate of an individual utterance was slowed, thus replicating Dilley and Pitt's results. In addition, listeners were less likely to perceive a function word when global speech rate was slower, suggesting that the disappearing-word effect was also influenced by the rate of utterances across the whole experiment. Together, these results do not only confirm that entrainment affects listeners' timing predictions, but also that such predictions are based on integrating entrainment that takes place over different timescales: both over the course of a speaker's individual turn, and over many turns.

Studies demonstrating a disappearing word effect suggest that listeners entrain to their interlocutor's syllable rate (both over a single utterance and over many utterances) and predict that the rate of forthcoming syllables will continue in line with the entrained rate. However, we do not know whether these predictions (made during language comprehension) can affect the timing of articulation (during language production), as suggested by Garrod and Pickering (2015). To investigate this issue we presented participants with simple questions (e.g., *Do you have a dog?*) and instructed them to answer either *yes* or *no*. To determine whether entrainment over multiple timescales influences when listeners launch articulation of their response (i.e., the duration of inter-turn intervals), we used a method similar to Pitt et al. (2016; see also Dilley & Pitt, 2010) and manipulated both the context (e.g., *Do you have a…*) and final word (e.g., *dog*?) rate. But unlike Pitt et al., who presented contexts at a natural of slow rate, we manipulated each component, so that they were presented at either a natural rate (normal spoken rate) or a speeded rate (compressed by a factor of 0.5, so it was twice as fast as its natural rate).

If comprehension and production share timing mechanisms, then we expect the timing of articulation to be influenced by the speech rate of both the context and the final word of the speaker's question, in a manner consistent with research demonstrating that comprehenders make timing predictions based on entrainment over multiple timescales (e.g., Baese-Berk et al., 2014). First, we expect listeners to entrain to the context rate of the speaker's utterance, which will in turn lead them to predict that the speaker will produce their final syllable at the same rate. Thus, listeners should respond later after contexts produced at a slower than a faster rate because they predict that the speaker will reach the end of their final syllable later (consistent with predictive entrainment as demonstrated by Dilley & Pitt, 2010). But additionally, listeners should adjust their timing predictions with each new syllable that they listen to; this would be generally consistent with entrainment over multiple timescales (Baese-Berk et al., 2014) and also specifically consistent with accounts that suggest listeners adjust the phase of their entrainment on a syllable-by-syllable basis (e.g., Giraud & Poeppel, 2012; Peelle & Davis, 2013). This means that listeners will respond earlier when the final word of the speaker's turn is produced at a faster than a slower rate because, upon encountering a fast final syllable they adjust their prediction so that they now predict that the turn will end earlier than they had expected on the basis of context alone. We term this the *tightly yoked account*, since it assumes that the timing mechanisms in comprehension can immediately affect language production.

However, it is also possible that comprehension and production share timing mechanisms, but changes in these representations during comprehension (for example, when the speaker suddenly changes their rate of syllable production) do not immediately affect language production. Previous research showing that interlocutors entrain on inter-turn intervals (e.g., Street, 1984) has kept the rate of utterances fairly consistent throughout turns, and so it is unclear how quickly changes in rate during comprehension affect the timing of subsequent production. Moreover, although there is some indication that entrainment to speech rate observed in auditory areas may be linked to brain activation in areas involved in the production of speech (e.g., Park et al., 2015), it is unclear whether timing in comprehension affects timing in production directly, or rather via an indirect mechanism. If this *loosely yoked account* is correct, then response times should be influenced by the context rate of the utterance, such that participants should respond later when the context is spoken at a natural rate rather than at a speeded rate, because there is time for the indirect mechanism to affect the initiation of articulation; but it is instead unlikely that response times would be strongly influenced by the rate of the final word.

Of course, another possibility is that comprehension and production do not share timing mechanisms (*separate mechanisms account*). If this is the case, then we expect no effects of context or final word rate on the timing of participants' responses. This account appears unlikely, however, since previous studies demonstrating speech rate convergence (e.g., Jungers & Hupp, 2009) suggest that entrainment during comprehension can influence the speech rate of utterances during language production.

Although none of these accounts predict an interaction between context and final word rate, we included this interaction in our analyses to control for other factors that may affect our results. First, participants may be surprised to encounter a rate change, and so we might expect a smaller effect of final word rate when a rate change occurs and the context is natural (i.e., natural-speeded vs. natural-natural condition) rather than speeded (speeded-natural vs. speeded-speeded condition). When a natural context is followed by a speeded final word (natural-speeded), listeners should respond more quickly than when the final word is also produced at a natural rate (natural-natural). However, this speeding effect will be counteracted by the slowing down effect of surprise (because a speeded final word comes after a rate change). When a speeded context is followed by a natural final word (speeded-natural), however, listeners should respond even later than would be expected based on entrainment to final word rate alone, because the delay in responses that we expect as a result of surprise adds to the delay that we expect after a natural final word (i.e., compared to speeded-speeded).

Furthermore, some research suggests that final lengthening can act as a turn-yielding cue (e.g., Gravano & Hirschberg, 2011). If listeners in our experiments use final lengthening as a cue to start articulation of their response, then they should respond earlier in the speeded-natural condition, in which the final word is lengthened in comparison to the rest of the utterance, than in the speeded-speeded condition. Final lengthening should not influence response times in the natural-natural and natural speeded conditions, thus leading to an interaction between final word and context rate.

## 3.2. Experiment 5

In Experiment 5, we tested the three accounts of how speech rate entrainment during comprehension can affect when listeners launch articulation of their response during conversational turn-taking. To do so, we used a verbal *yes/no* question-answering task and manipulated the speech rate of these questions using time compression, so that the context (e.g., *Do you have a…*) and final word (made up of a final syllable; e.g., *dog*?) were either compressed by a factor of 0.5 (i.e., twice as fast as the natural spoken rate; *speeded* conditions) or presented at the spoken rate (*natural* conditions). In other words, we created four conditions where a natural or speeded context was combined with a natural or speeded final word (natural-natural, natural-speeded, speeded-speeded, and natural-natural conditions).

Both the tightly and loosely yoked accounts predict a main effect of context rate, such that participants should respond earlier when context is speeded rather than natural. However, the accounts make different predictions regarding effects of final word rate. Since the loosely yoked account assumes that changes in timing representations in comprehension (i.e., after encountering a single syllable at a different rate at the end of a turn) do not immediately affect production, it predicts no effect of final word rate. The tightly yoked account, in contrast, predicts an effect of final word rate, such that listeners should respond earlier when the final word is speeded rather than natural. In addition, we tested for an interaction between these predictors to control for alternative factors (e.g., surprisal and final lengthening) that may affect response times.

### 3.2.1. Method

#### *3.2.1.1. Participants*

Thirty-two native English speakers (4 males; *M*age = 19.44) at the University of Edinburgh participated in exchange for course credit or £4. Participants had no known speaking, reading, or hearing impairments.

#### *3.2.1.2. Materials*

Participants listened to 124 questions. All questions were recorded by a native English male speaker, who was instructed to read the utterances as though “you are asking a question and expecting a response”. Since previous research suggests that prosodic cues play a role during turn-taking (e.g., Duncan, 1972), we inspected our audio recordings for such cues both auditorily and phonetically (i.e., waveform and spectrogram) using Praat (Boersma & Weenink, 2002). All questions had falling boundary tones. Boundary tone judgements were validated by a second coder, which resulted in a Cohen's kappa of 1. In addition, some research suggests that pitch downstep, which occurs when the pitch of each syllable is lower than the previous syllable (Beckman & Pierrehumbert, 1986), can act as a turn-yielding cue (Cutler & Pearson, 1985). Although two independent raters could not agree on downstep judgments for the stimuli, this disagreement should not pose a problem for later interpretation, given that the manipulation is within-items and time compression does not alter the pitch of utterances.

We manipulated the speech rate of these questions using a time compression factor of 0.5. Stimuli were time compressed using the Pitch-Synchronous Overlap and Add (PSOLA) algorithm in Praat (Moulines & Charpentier, 1990). This method altered utterance speech rate (so it was produced twice as fast, i.e., speeded utterances; see Table 5) but left the speech stream unaltered in the frequency-domain (preserving e.g., pitch and segmental information). Both the natural and speeded utterances were divided into a context and a final word (which included any pause prior to the onset of the final word) to create two versions of each (speeded context, natural context; speeded final word, natural final word). Context and final word regions were then recombined to create four stimuli conditions: (i) natural-natural, where both the context and the final word were presented at the spoken rate; (ii) natural-speeded, where the context was presented at the spoken rate, but the final word was compressed; (iii) speeded-speeded, where both the context and the final word were compressed; and (iv) speeded-natural, where the context was compressed, but the final word was presented at the spoken rate. Thus, speech rate either stayed the same throughout the questions or changed on the final word.

Table 5. The means (and standard deviations) of the total duration, context duration, and final word duration (ms) for the four stimuli conditions in Experiment 5.

| Context | Final | Total Duration | Context Duration | Final Word Duration |
|---|---|---|---|---|
| Natural | Natural | 1838 (416) | 1341 (418) | 497 (96) |
| | Speeded | 1591 (414) | 1341 (418) | 250 (48) |
| Speeded | Natural | 1166 (217) | 669 (208) | 497 (96) |
| | Speeded | 910 (218) | 669 (208) | 250 (48) |

To ensure our time compression manipulation did not make the sentences unintelligible (given that participants were expected to comprehend the questions before answering), we assessed intelligibility using a pre-test, in which 28 further participants (6 males; *M*age = 19.61) listened to the questions and typed exactly what they heard the speaker say. We calculated the average intelligibility of each utterance by comparing the number of words in the question to the number of words participants correctly identified. Any obvious spelling mistakes or typing errors (i.e., from keys around the target letter or missing letters) were scored as correct, but morphological mismatches were not (e.g., *younger* would be scored incorrect if the target was *young*; Davis et al., 2005; Loebach, Pisoni, & Svirsky, 2010). Although an ANOVA showed that intelligibility was lower in the speeded than the natural context conditions ($p$ = .01; all other comparisons $p$ > .05), it was high (> 98%) in all conditions (mean of 99.6% in the natural-natural condition, 99.2% in the natural-speeded condition, 99.1% in the speeded-natural condition, and 98.9% in the speeded-speeded condition). Moreover, if intelligibility influences answer times, then we would expect participants to answer later in the speeded context conditions (where intelligibility is lower) than the natural context conditions (i.e., the opposite of what both the tightly and loosely yoked accounts predict).

Previous work indicates that listeners may use content predictions to prepare (Corps, Crossley, Gambi, & Pickering, 2018). To limit between-items variability, we selected only questions that were unpredictable in content. We assessed the predictability of our stimuli using a cloze pre-test, in which 21 further participants from the same population (3 males, *M*age = 21.43) were presented with the questions (missing their final word) and were instructed to "complete each fragment with the word or words that you think are most likely to follow the preceding context of the question." (i.e., we used a cloze task; Taylor, 1953). The content predictability of fragments was assessed using Shannon entropy (i.e., $-\Sigma p_i \log_2(p_i)$, where $p_i$ is the proportion of times each completion occurred for a given fragment; C. E. Shannon, 1948), which is low (a minimum of 0) when completions are the same across participants (i.e., content is predictable), and high (a maximum of 4.39 when each of the 21 participants in the pre-test provided a different completion) when completions are different. Content entropy was low (see Table 6), indicating that questions were unpredictable and did not predict a particular continuation. In addition, we used cloze probability (Taylor, 1953) to calculate the percentage of participants who provided a particular continuation. All final words we selected had low cloze probability (see Table 6).

Note that using data from the same pre-test we were also able to check that question fragments did not differ in length predictability (the number of words that participants would expect to complete them). We calculated length predictability using entropy (using the same formula for content entropy, but *pi* is the proportion of times each completion length occurs for a given fragment), which was low for all fragments (see Table 6). Furthermore, all questions were completed with a single word by at least 70% of participants. Thus, one word completions were predictable for all our stimuli, and differences in the length predictability of questions could not confound our results (see Magyari & De Ruiter, 2012). All completions consisted of a single monosyllabic word, to ensure that the final word of all stimuli provided participants with the same amount of information (i.e., a single syllable) about a change of rate.

Table 6. The means (*M*) and standard deviations (*SD*s) of our measures of content predictability, length predictability, difficulty, and plausibility for stimuli in Experiment 5.

| | *M* | *SD* |
|---|---|---|
| Completion Length Entropy[a] | 0.63 | 0.39 |
| Completion Length Cloze[b] | 86% | 9% |
| Completion Content Cloze[c] | 6% | 3% |
| Question Fragment Entropy[d] | 3.28 | 0.63 |
| Question Difficulty[e] | 6.16 | 0.05 |
| Question Plausibility[e] | 5.82 | 0.08 |

[a] Entropy of the length (in number of words) of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion length.

[b] Percentage of participants who provided the word length of the selected completion used in the main experiment (always a single word) as a continuation in the cloze task.

[c] Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

[d] Entropy of the content of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.

[e] Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, and 7 indicated that the question was very plausible/easy to answer.

Finally, we measured the difficulty and plausibility (see Table 6) of all questions using ratings during a second pre-test, in which 12 further participants (6 males; *M*age = 29.92) rated (i) how difficult they would find it to answer the question if asked, and (ii) whether the question made sense. Both ratings were made on a scale of 1 (very implausible/difficult) to 7 (very plausible/easy). The mean ratings of 6.16 for difficulty and 5.82 for plausibility indicated that the questions were judged to be fairly easy and plausible.

### *3.2.1.3. Design*

Both context rate (speeded vs. natural) and final rate (speeded vs. natural) were varied within participants and items, and so there were four versions of each stimulus. We created four experimental lists (each containing 124 questions) using a Latin Square procedure, so that all participants saw one version of each item and 31 items from each condition.

### *3.2.1.4. Procedure*

Stimulus presentation and data recording were controlled using E-Prime (version 2.0). A fixation cross (+) appeared 500 ms before question onset, and the screen turned red as audio playback began. Participants pressed a button on the response box to start audio playback of the question, and were told to: "Answer either *yes* or *no* as quickly as possible. Do not wait until the speaker has finished the question and has stopped speaking. Instead you should answer as soon as you expect the speaker to finish." Thus, participants were encouraged to respond as quickly as possible. Participants responded using the microphone provided, and playback stopped as soon as a voicekey response was recorded.

At the start of the experiment, participants completed twelve practice trials (three from each of the four conditions) to familiarize themselves with the experimental procedure. The 124 stimuli were individually randomized for each participant, and participants were given the opportunity to take a break after every 31 items.

### 3.2.2. Data Analysis

Of the 3968 answers, 175 (4.41%) were discarded because the audio recording was unclear and so the answer could not be categorized as either *yes* or *no*. We removed a further three (0.08%) answer times greater than 10000 ms because they were clear outliers. We then replaced any responses falling 2.5 standard deviations above (80; 2.02%) or below (27; 0.68%) the by-participant mean answer time with the respective cut-off value.

We first calculated answer times from final word offset (i.e., question offset), as this measure is equivalent to inter-turn intervals in natural dialogues. However, our primary dependent variable was response time measured from final word onset (which was derived by adding final word duration to the participant's response time as measured from question offset). We assume that participants prepared their own response after the onset of the final word of the speaker's utterance, because all of our questions were unpredictable in content and participants could determine what to respond only after the speaker began producing the critical final word. But since participants prepared their response after the onset of the final word, then it also means that they had more time available for response preparation and initiating articulation when the final word was longer; thus, we would expect them to respond closer to final word offset when the final word was longer. Indeed, there was a negative correlation between final word duration and answer time from question offset ($r$ = -0.21, $p$ < .001), such that questions with longer final words tended to elicit earlier responses than those with shorter final words.

As a consequence, analyses from final word offset may not be informative about entrainment, as participants may respond closer to the offset of natural than speeded final words simply because natural words are longer, which gives them more time to prepare an answer. If this preparation advantage following natural words is sufficiently large, it may even mask any effect of final word rate. Instead, analyses from word onset are not confounded by response preparation and thus provide a better index of entrainment.

To check our assumptions, as well as testing our hypotheses about entrainment, we evaluated the effects of context rate and final rate on answer times from both final word onset and offset with linear mixed effects models (LMM; Baayen et al., 2008) using the *lmer* function of the *lme4* package (version 1.1-12; Bates et al., 2015) in RStudio (version 0.99.896) with a Gaussian link function. In all instances, we fitted models using the maximal random effects structure justified by our design (Barr et al., 2013) but correlations among random effects were fixed to zero to aid model convergence (Matuschek et al., 2017). We fitted the full model, in which answer speed (from either final word onset or offset) was predicted by Context Rate (reference level: natural vs. speeded), Final Word Rate (reference level: natural vs. speeded), and their interaction.

To account for other factors that may affect answer times, we also included three further predictors as main effects. Participants tend to answer *yes* more quickly than *no* (e.g., Strömbergsson et al., 2013), and so we also included Answer (reference level: no vs. yes) as a predictor. In addition, we included Answer Agreement, which is the absolute difference between the percentage of participants who answered *yes* and the percentage who answered *no* (i.e., with 100 occurring if all answered *yes* or all answered *no*, and 0 occurring if half answered each way), as participants answer more quickly when Answer Agreement is higher (Corps et al., 2018). Finally, some studies have found a negative relationship between duration of the whole turn (not just of the final word) and response times (e.g., De Ruiter et al., 2006), and so we also included Question Duration as a predictor.

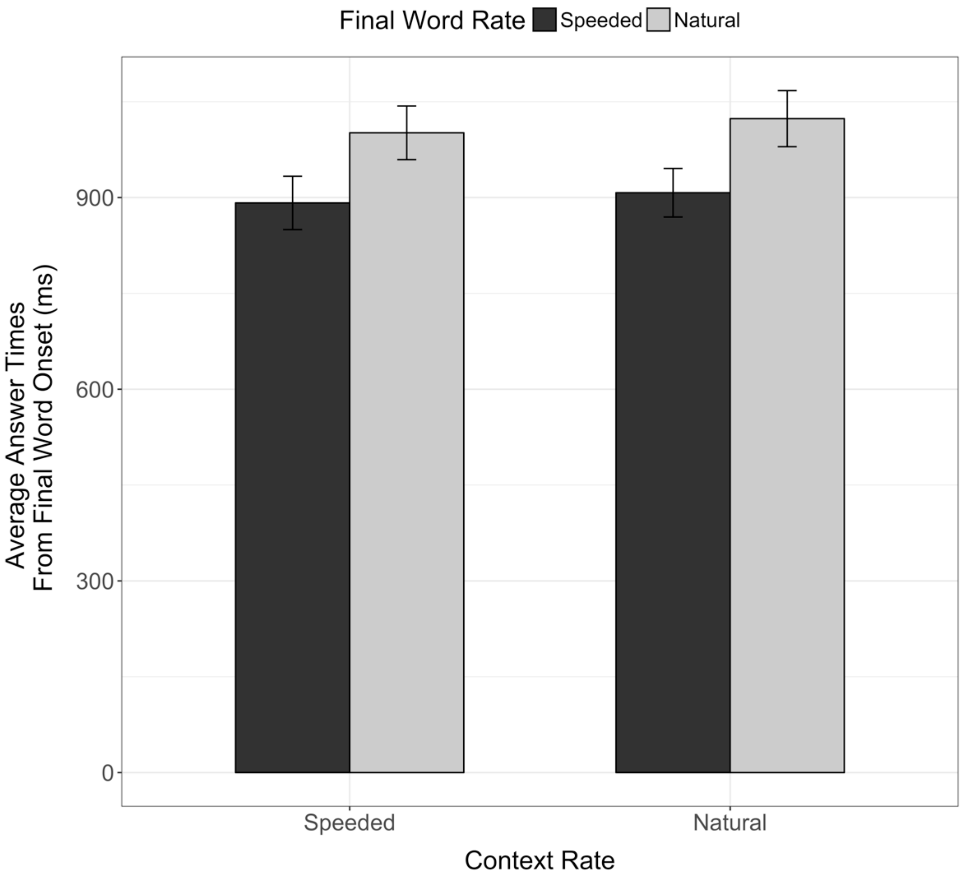
All predictors were contrast coded (-0.5, 0.5; where relevant) and centered before being added to the model. We assume that a *t* value of greater than 1.96 indicates significance at the 0.05 alpha level (Baayen et al., 2008), and we report coefficient estimates (*b*), standard errors (*SE*), and *t* values for each predictor.

### 3.2.3. Results

#### *3.2.3.1. Analysis from final word onset: Rate entrainment*

Consistent with both the tightly and loosely yoked accounts, we found an effect of Context Rate: Participants answered earlier after a speeded than a natural context (*b* = -42.17, *SE* = 18.90, *t* = -2.23; mean answer times for speeded = 947 ms vs. natural = 966 ms; see Fig. 10). But in addition, participants answered earlier when the final word was speeded than when it was natural (*b* = -122.38, *SE* = 13.77, *t* = -8.89; mean answer times for speeded = 899 ms vs. natural = 1012 ms), consistent with the tightly yoked account but not with the loosely yoked account.

Figure 10. Observed means of answer times (ms) from final word onset for the four conditions in Experiment 5. Error bars represent ± 1 standard error from the mean.



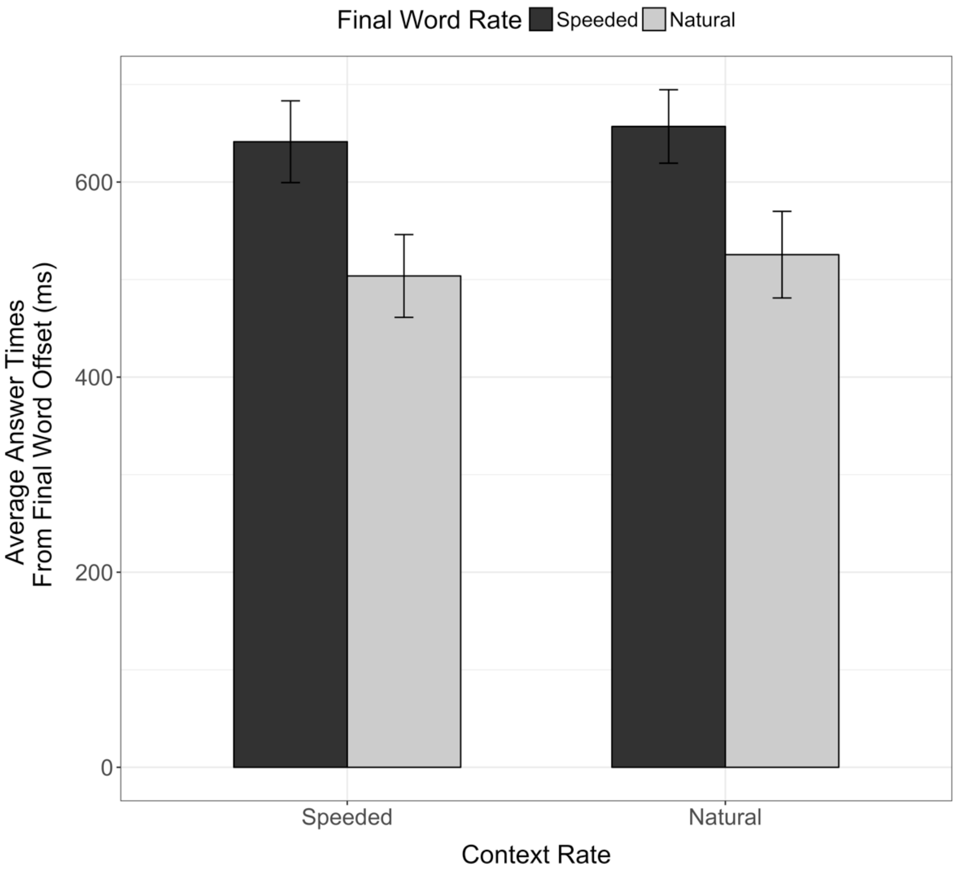
There was no interaction between Context Rate and Final Word Rate (*b* = 8.60, *SE* = 18.91, *t* = 0.46), clearly ruling out the possibility that answer times were driven by intelligibility, final word lengthening, or surprise at a rate change. First, if answer times were driven by intelligibility, then we would have expected participants to be slower to answer in the speeded context conditions, where intelligibility was lower, but instead they were slower in the natural context conditions. Second, if answer times were driven by final lengthening, then we would have expected participants to answer earlier in the speeded-natural conditions than all other conditions because the final word was lengthened in comparison to the rest of the utterance. Finally, if answer times were driven by surprise, then we would have expected a larger effect of Final Word Rate after speeded than natural contexts.

We also found an effect of Answer Agreement (*b* = -27.77, *SE* = 8.70, *t* = -3.19): Listeners responded earlier when Agreement was higher. Furthermore, participants were quicker to answer *yes* than *no* (*b* = -73.90, *SE* = 10.18, *t* = -7.26; mean answer times for yes = 915ms vs. no = 998ms). In contrast, there was no effect of Question Duration (*b* = -17.14, *SE* = 11.62, *t* = -1.47).

### *3.2.3.2. Analysis from final word offset: Response preparation*

As in the analysis from final word onset, we found that participants answered earlier when context was speeded rather than natural (*b* = 65.54, *SE* = 18.00, *t* = -3.64; mean answer times for speeded = 572 ms vs. natural = 590 ms; see Fig. 11). However, the effect of Final Word Rate was in the opposite direction to that in the analysis from final word onset: Participants answered earlier after a natural than a speeded final word (*b* = 116.32, *SE* = 13.84, *t* = 8.40; mean answer times for natural = 514 ms vs. natural = 649 ms). As we discussed in the Data Analysis section, this effect most likely occurred because a slow final word gives participants more time to prepare their own verbal response. Therefore, this finding is not informative as to whether the listener adjusted their timing predictions after the rate change, but rather it shows that preparation time has a large effect on the duration of inter-turn intervals. There was no interaction between Context Rate and Final Word Rate (*b* = 8.42, *SE* = 19.11, *t* = 0.44).

Figure 11. Observed means of answer times (ms) from final word offset for the four conditions in Experiment 5. Error bars represent ± 1 standard error from the mean.



Again, both Answer Agreement (*b* = -39.40, *SE* = 7.91, *t* = -4.98) and Answer Type (*b* = -73.81, *SE* = 10.06, *t* = -7.34; mean answer times for yes = 541ms vs. no = 620ms) were predictors of answer times. But, unlike the analysis from final word onset, Question Duration was a negative predictor: Participants answered earlier when questions were longer in duration (*b* = -34.16, *SE* = 10.88, *t* = -3.14).

### 3.2.4. Discussion

In Experiment 5, we investigated how speech rate entrainment during comprehension influenced the timing of response articulation during language production (e.g., Garrod & Pickering, 2015; Wilson & Wilson, 2005) using a *yes/no* question-answering task. We manipulated the context (e.g., *Do you have a…*) and final word (e.g., *dog*?) rate of our questions, so that each component was either produced at a natural or a speeded rate. Results from final word offset were likely affected by response preparation time. The results from final word onset, which were instead unaffected by response preparation, were thus crucial for testing our hypotheses about rate entrainment.

Consistent with both the tightly and loosely yoked accounts, we found that participants answered earlier after questions with a speeded rather than a natural context rate, suggesting that they entrained to the context rate (i.e., over multiple syllables) of the speaker's question. Consistent with the tightly yoked, but not loosely yoked, account, we also found that listeners answered earlier when the final word was speeded (speeded-speeded and natural-speeded conditions) rather than natural (natural-natural and speeded-natural conditions), suggesting that listeners adjusted their timing predictions immediately after encountering a final syllable that differed in rate from the question, and these predictions immediately affected the timing of subsequent production. Taken together, our findings are consistent with the tightly yoked account, and suggest that comprehension and production share timing representations.

Interestingly, question duration predicted answer times when they were measured from final word offset (as in previous studies that reported effects of question duration), but not when they were measured from final word onset. This suggests that the length of the final word is an important contributor to the question duration effect observed in previous question-answering studies (Corps et al., 2018), perhaps because such effect is linked in large part to the amount of time available for response preparation, and response preparation did not take place until after the onset of the final word in our materials (as they were all unpredictable).

In addition, answer times from final word offset showed that participants answered earlier when the final word was natural (and therefore longer in duration) rather than speeded (and therefore shorter in duration), which further confirms our assumption that participants began response preparation while listening to the speaker's final word. However, answer times from final word onset showed that participants answered earlier when this word was speeded rather than natural. In other words, our final word effect was reversed when we analyzed from final word onset compared to when we analyzed from final word offset. Such reversal of the effect across analyses suggests that the response preparation advantage afforded by natural final words is so large that it can mask the effect of adjusting to final word rate.

However, the effect of context rate did not depend on analysis location in the same way. This is potentially worrying, as it may suggest that the final rate finding is not due to adjusting of speech entrainment after all, or else it should behave similarly to the context rate effects. Indeed, there is an alternative explanation for the final rate findings. Perhaps listeners respond closer to final word onset when this word is speeded because speeded words are recognized earlier and this in turn allows them to start response preparation earlier. In other words, it is possible that our final rate manipulation affects response times because it affects when response preparation can start rather than because it affects entrainment. To test whether this is the case, we crossed final rate with a manipulation of content predictability in Experiment 6. By making the final word predictable in half the questions, we allow participants to start preparation before a rate change occurs and before they even hear the final word. Thus, if the final rate effect is indeed due to easier recognition, we should find that it is reduced when the final word is predictable.

## 3.3. Experiment 6

In Experiment 5, we found that listeners' timing of articulation entrained to the rate at which the speaker had produced the majority of their question (i.e., the context rate). Additionally, we found that listeners entrained to the rate of the speaker's final word, and launched articulation in line with this entrainment. Although we interpret these findings as consistent with a tightly yoked account, these results are also consistent with the possibility that our speech rate manipulation affected response preparation. Specifically, listeners may respond closer to the onset of speeded final words because they can recognize them earlier, and can thus begin response preparation earlier.

In Experiment 6, we tested this alternative explanation by varying whether response preparation was possible only after recognizing the final word (unpredictable questions; e.g., *At University, do you study maths*?), or was possible before hearing this word (predictable questions; e.g., *Are dogs your favorite animal*?). Indeed, previous research suggests that listeners prepare earlier when the content of the speaker's final word is predictable (Corps et al., 2018). Final words in the predictable condition were always consistent with participants' predictions based on context. We also varied the syllable length of the final word of our stimuli to generalize the final word effect to multi-syllable items. If the results of Experiment 5 are due to response preparation and not rapid adjusting of timing representations, then we would expect the final word effect to be reduced when content is predictable, and participants can begin preparation before the speaker's final word, compared to when content is unpredictable, and participants can begin preparation only on the speaker's final word. In other words, we expect an interaction between content predictability and final word rate.

If, however, our results are due to rapid adjusting (as predicted by the tightly yoked account), then we expect the effect of final word rate to be the same, regardless of the predictability of the final word of the speaker's question. In other words, it should not matter whether preparation can occur early (as in the predictable condition) or not (as in the unpredictable condition): The effect of final word rate should be similar in both cases, because the speech rate manipulation should affect the timing of response articulation via entrainment, but not the timing of preparation. Thus, we expect to replicate previous research and find an effect of content predictability, such that listeners should respond earlier when content is predictable than unpredictable (e.g., Corps et al., 2018), but crucially we would not expect an interaction between content predictability and final word rate. Of course, we would also expect to replicate the final word effect from Experiment 5, and find that listeners respond closer to the onset of a speeded than a natural final word. Note that finding no interaction between content predictability and final word rate would also be consistent with accounts that suggest preparing a response and timing its articulation are independent processes, controlled by separate mechanisms (e.g., Levinson & Torreira's (2015) early-planning account).

### 3.3.1. Method

#### *3.3.1.1. Participants*

Thirty-two new participants from the same population as in Experiment 5 (9 males; *M*age = 20.10) took part on the same terms.

#### *3.3.1.2. Materials*

Using the same norming procedure as in Experiment 5 (22 native English speakers; 6 males; *M*age = 18.5), we elicited completions for 292 question fragments. We assessed the length and content predictability of stimuli and completions using the same procedure as in Experiment 5, but rather than selecting only unpredictable questions, we instead selected 35 predictable content and 35 unpredictable content questions (70 stimuli in total). As intended, stimuli in the predictable content condition had significantly higher content entropy and cloze than those in the unpredictable condition (all *p*s < .001; see Table 7). The two conditions were matched for average length entropy (*p* > .17), completion length occurrence (*p* > .46), difficulty, and plausibility (all *p*s > .17; pre-tested with 12 participants, 4 males, *M*age = 18.5).

Table 7. The means (and standard deviations) of content predictability, length predictability, difficulty, plausibility, and intelligibility for stimuli in the predictable and unpredictable conditions of Experiment 6.

| | Predictable | Unpredictable |
|---|---|---|
| Completion Length Entropy[a] | 0.09 (0.18) | 0.14 (0.17) |
| Completion Length Cloze[b] | 98% (5%) | 98% (3%) |
| Completion Content Cloze[c] | 91% (9%) | 6% (3%) |
| Question Fragment Entropy[d] | 0.46 (0.45) | 3.09 (0.47) |
| Question Difficulty[e] | 6.48 (0.62) | 6.35 (0.49) |
| Question Plausibility[e] | 6.69 (0.39) | 6.65 (0.32) |

[a] Entropy of the length (in number of words) of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion length.

[b] Percentage of participants who provided the word length of the selected completion used in the main experiment (a single word in all conditions) as a continuation in the cloze task.

[c] Cloze percentages of the selected completion. If cloze percentage is higher, then participants converged on a completion.

[d] Entropy of the content of question fragments presented to participants in the cloze task. If entropy is lower, then participants converged on a completion.

[e] Difficulty and plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible/difficult to answer, while 7 indicated that the question was very plausible/easy to answer.

Although we varied the syllable length of our stimuli to determine whether our final word effect generalized to multi-syllable items, we also made sure that the two predictability conditions had the same numbers of one (14), two (13), and three (8) syllable completions. All questions were recorded using the same procedure as in Experiment 5. As two of our manipulations were between items (final syllable length and content predictability), it was important to check whether the conditions differed acoustically in any systematic way. Ten (14%) of the utterances had creaky voice (four in the predictable condition; six in the unpredictable condition). In most cases the stressed syllable was the first syllable of the word (91% in the predictable condition; 94% in the unpredictable condition). As in Experiment 5, all questions had falling boundary tones and 61 of the utterances had a downstep in pitch (89% in the predictable condition; 86% in the unpredictable condition). Both judgments were again validated by the same second coder as in Experiment 5, who rated 25% of the stimuli, and this time was in perfect agreement with the first coder (Cohen's kappa = 1 for both boundary tone and downstep judgments).

Using the same time-compression method as in Experiment 5, we manipulated the rate of the final word of each question (either natural or speeded; see Table 8) and created two versions of each stimulus (natural-natural and natural-speeded). Importantly, there were no interactions between content predictability, final word condition, and syllable length ($p$ = .73) and no two-way interactions (all $p$s > .15), suggesting that the rate manipulation was comparable across predictability conditions and the different syllable lengths. All conditions were matched for average intelligibility (all *ps* > .90; mean of 99.9% in all conditions) using the same procedure as Experiment 5.

Table 8. The means (and standard deviations) of the context and final word durations (ms) of questions in Experiment 6. The final column provides the difference in the means of the final word durations of the natural and speeded final words.

| Content | Final Word Syllable Length | Context Duration | Natural Final Word Duration | Speeded Final Word Duration | Difference in Final Word Duration |
|---|---|---|---|---|---|
| Predictable | 1 | 1699 (608) | 414 (89) | 207 (45) | 207 |
| | 2 | 1621 (423) | 466 (107) | 233 (53) | 233 |
| | 3 | 1744 (696) | 482 (80) | 241 (40) | 241 |
| Unpredictable | 1 | 1250 (361) | 442 (132) | 221 (66) | 221 |
| | 2 | 1183 (292) | 515 (83) | 257 (42) | 257 |
| | 3 | 1260 (474) | 558 (66) | 279 (33) | 279 |

#### *3.3.1.3. Design*

Predictability (predictable vs. unpredictable) was manipulated within participants but between items. Final condition (speeded vs. natural) was manipulated within both participants and items. We created two stimulus lists (each containing 70 stimuli) using a Latin Square procedure, such that each list contained: (i) 35 predictable and 35 unpredictable stimuli; and (ii) 14 one syllable completions, 13 two syllable completions, and 8 three syllable completions from each of the predictable and unpredictable conditions. Every combination of predictability and final rate condition occurred once across these two lists.

#### *3.3.1.4. Procedure*

The procedure was identical to Experiment 5, except that participants completed 12 practice trials (1 from each of the four conditions; one single syllable completion, one two syllable completion, and one three syllable completion) and were given the opportunity to take a break after the first 35 stimuli.

### 3.3.2. Data Analysis

Answer times were calculated using the same procedure as Experiment 5. In the unpredictable content conditions, we again expected participants to prepare after final word onset, so that response times measured from final word offset could have been affected by response preparation. In contrast, participants could (in principle) prepare a response before final word onset in the predictable content conditions. But since we did not manipulate the duration of the context in Experiment 6, the amount of time available for preparation before the final word could not affect answer times measured from final word onset.

We discarded 75 (3.35%) of the 2240 responses because the audio recording was unclear and so the answer could not be categorized as either *yes* or *no*. We then discarded a further three (0.13%) answer times greater than 10000 ms, and replaced 33 (1.47%) answer times at the upper limit and 21 (0.94%) at the lower limit. All data analyses, methods, and predictors were identical to those used in Experiment 5.

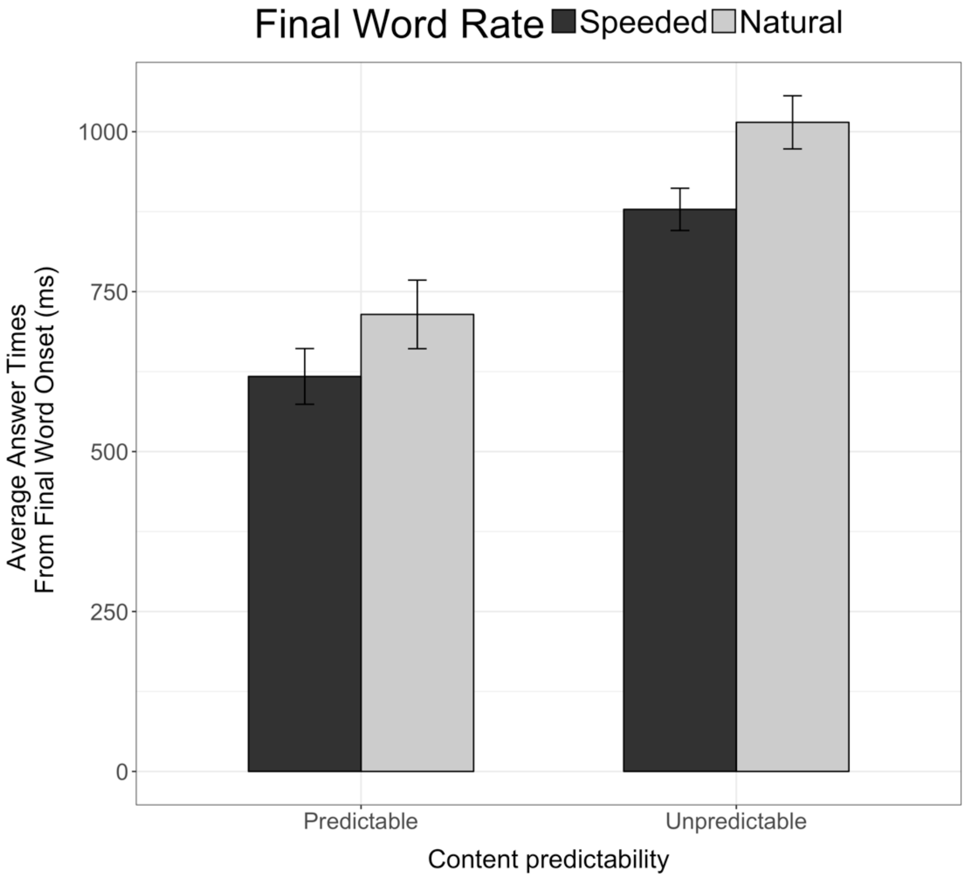
Thus, we again fitted the full model, in which answer times (from final word onset or offset) were predicted by Content Predictability (reference level: unpredictable vs. predictable), Final Word Rate (reference level: speeded vs. natural), and their interaction. Since the loosely yoked account suggests changes in speech rate during comprehension may not immediately affect subsequent production, we also included the number of syllables of the final word (and its interactions) as a continuous predictor to determine whether there was an interaction with Content predictability and Final word rate. We again included all three control variables (Answer, Answer Agreement, and Question Duration) from Experiment 5 to account for possible confounding factors.

### 3.3.3. Results and Discussion

#### *3.3.3.1. Analysis from final word onset: Rate entrainment*

Participants responded earlier when content was predictable than unpredictable (*b* = -201.81, *SE* = 41.18, *t* = -4.90; mean answer times for predictable = 665 ms vs. unpredictable = 947 ms; see Fig. 12), suggesting that listeners were sensitive to the content predictability of the speaker's question and used this information to prepare a response as early as possible (e.g., Bögels et al., 2015; Corps et al., 2018). We also found a significant effect of Final Word Rate: Participants responded earlier after a speeded than a natural final word (*b* = -125.79, *SE* = 19.56, *t* = -6.43; mean answer times for speeded = 748 ms vs. natural = 865 ms), thus replicating Experiment 5. Finally, there was no interaction between Content Predictability and Final Word Rate (*b* = 38.50, *SE* = 30.54, *t* = 1.26), suggesting that our effect of final word rate in Experiment 5 did not occur simply because participants recognized the speaker's final word and began preparation earlier in the speeded than the natural condition. Instead, entrainment affected only response articulation, even after a single syllable differing in rate, consistent with the tightly yoked account. Accordingly, the number of syllables did not influence response times ($b$ = 15.59, $SE$ = 14.51, $t$ = 1.07) and did not interact with Final word rate ($b$ = -23.83, $SE$ = 15.71, $t$ = -1.52; all other comparisons $t$ < -1.52). Thus, our final word effect from Experiment 5 generalized to multi-syllabic words.

Figure 12. Observed means of answer times (ms) from final word onset for the four conditions in Experiment 6. Error bars represent ± 1 standard error from the mean.

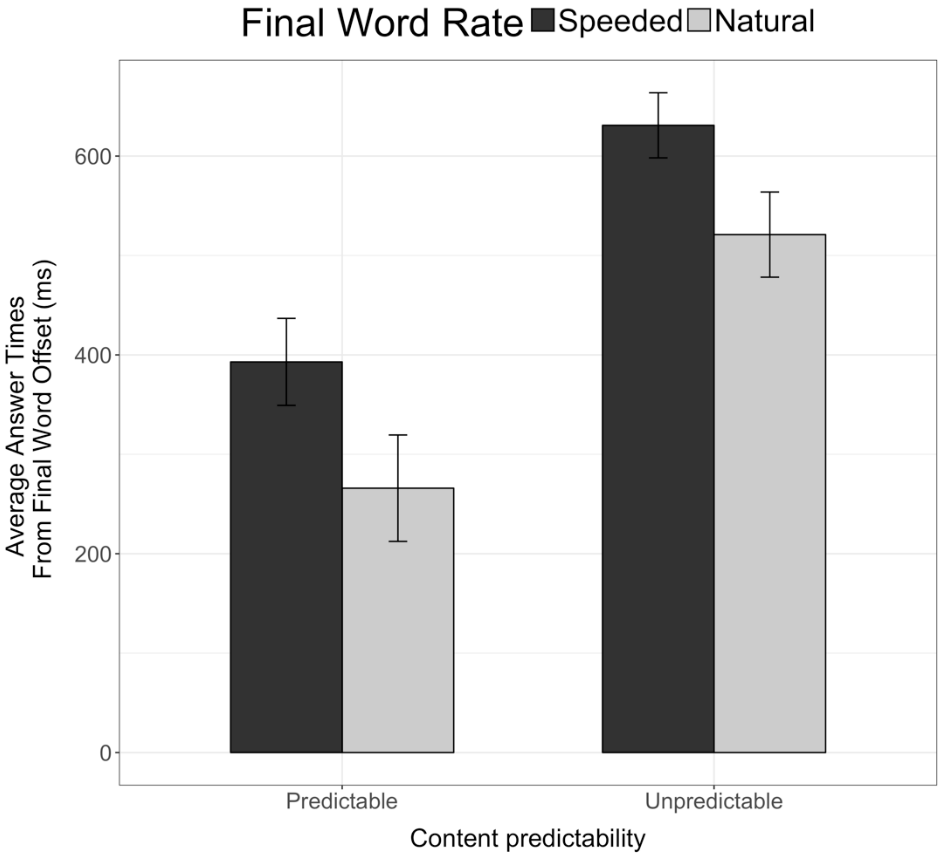


As in Experiment 5, Answer Agreement was a negative predictor of answer times (*b* = -46.51, *SE* = 15.89, *t* = -2.93), and participants were quicker to answer *yes* than *no* (*b* = -92.00, *SE* = 21.44, *t* = -4.29; mean answer times for yes = 733ms vs. no = 961ms). However, Question Duration again did not predict answer times (*b* = -25.44, *SE* = 16.01, *t* = -1.59).

### *3.3.3.2. Analysis from final word offset: Response preparation*

In our analysis from final word offset, we replicated the finding that that participants answered earlier when questions were predictable rather than unpredictable in content (*b* = -153.46, *SE* = 38.03, *t* = -4.04; mean answer times for predictable = 328 ms vs. unpredictable = 577 ms; Fig. 13). As in our analysis from final word offset in Experiment 5, participants answered earlier when the final word was natural rather than speeded (*b* = 106.70, *SE* = 19.41, *t* = 5.50; mean answer times for natural = 393 ms vs. speeded = 511 ms). Crucially, however, there was no interaction between these two factors (*b* = 15.92, *SE* = 30.45, *t* = 0.52). In addition, the number of syllables in the final word was not a significant predictor (*b* = -13.99, *SE* = 12.71, *t* = -1.01), and there was no two-way interaction between the number of syllables and Content Predictability (*b* = -20.36, *SE* = 24.50, *SE* = -0.80), and no three-way interaction between number of syllables, Content Predictability, and Final Word Rate (*b* = 36.36, *SE* = 30.57, *t* = 1.19).

Figure 13. Observed means of answer times (ms) from final word offset for the four conditions in Experiment 6. Error bars represent ± 1 standard error from the mean.



Answer Agreement was a negative predictor of response times (*b* = -59.56, *SE* = 13.99, *t* = -4.26) and participants answered *yes* faster than *no* (*b* = -90.29, *SE* = 20.95, *t* = -4.31; mean answer times for yes = 387ms vs. no = 585ms). Finally, Question Duration was a negative predictor (*b* = -31.70, *SE* = 14.04, *t* = -2.26), such that questions longer in duration elicited earlier answers than those shorter in duration. Together with the final word onset results, this effect replicates Experiment 5 and suggests that effects of question duration may be largely attributed to the final word.

## 3.4. General Discussion

In two experiments, we used a verbal question-answering task to investigate whether listeners time response articulation by entraining to a speaker's speech rate. Specifically, we investigated how tightly yoked timing representations during comprehension are to timing representations during subsequent production. We contrasted three accounts: (1) a tightly yoked account, which suggests that comprehension and production share timing representations, such that entrainment over multiple timescales (i.e., a single utterance and a single syllable) during language comprehension can immediately affect the timing of response articulation during language production; (2) a loosely yoked account, which suggests that comprehension and production share timing representations, but changes to speech rate entrainment during language comprehension (e.g., when the speaker suddenly changes their rate of syllable production) do not immediately affect language production; and (3) a separate mechanisms account, which suggests that comprehension and production do not share timing mechanisms, and so entrainment during language comprehension should not influence the timing of response articulation during language production. To distinguish these three accounts, we manipulated the speech rate of questions, so that a natural or speeded context was combined with a natural or speeded final word (in Experiment 5; in Experiment 6, the context was always natural and only the final word rate was manipulated).

In Experiment 5, we found that participants entrained to the context rate of the speaker's turn: They answered earlier when the context was speeded (twice as fast as its original rate) rather than natural. This context effect is consistent with evidence of speech-rate priming (e.g., Jungers & Hupp, 2009), but extends these findings and suggests that the rate of the speaker's turn influenced not just the rate of the listener's own response, but also the timing of its initiation. In addition to this context effect, we also found that listeners responded earlier (when measuring from final word onset) when the speaker's final syllable was speeded rather than natural, regardless of context rate. These results are consistent with a tightly yoked account: Listeners adjusted their entrainment after encountering a single syllable that differed in rate from the preceding context, and this entrainment during language comprehension then immediately affected the timing of response articulation during subsequent production.

In Experiment 6, we replicated this final word effect with final words of different syllable lengths. In addition, we found that participants responded earlier when the content of the final word of the speaker's question was predictable (e.g., *Are dogs your favorite animal*?) rather than unpredictable (e.g., *At University, do you study maths*?), suggesting that they used these predictions to prepare a verbal response (e.g., Bögels et al., 2015; Corps et al., 2018). However, content predictability did not influence the effect of final word rate. This ruled out the possibility that participants in Experiment 5 responded earlier when the final word was speeded rather than natural because the disambiguating information necessary for recognizing the speaker's final word occurred earlier (and subsequent response preparation could also occur earlier).

Note that this lack of interaction also rules out the possibility that the final word effect in Experiment 6 occurred because even when participants could predict the final word (i.e., in the predictable conditions), they still waited until they could verify their prediction before launching articulation of their response. Although this "prediction check" would likely be quicker when the final word was speeded rather than natural, it is likely that it proceeded on the basis of partial acoustic information (e.g., Dahan & Tanenhaus, 2004; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Crucially, it appears that when listeners cannot predict the speaker's final word (i.e., in the unpredictable conditions), they need to process more of the final word in order to recognize it, compared to listeners who can predict the final word, as we find a clear effect of content predictability. Thus, if listeners are indeed carrying out a prediction check, then the effect of final word rate on recognition should still have been smaller in the predictable than in the unpredictable conditions, because the shorter the portion of the word that is checked, the less scope there is for the final word rate manipulation to make a difference (e.g., compare being able to launch articulation after 50 vs. 100 ms to after 300 vs. 600 ms).

Together, these findings are consistent with previous studies (e.g., Baese-Berk et al., 2014) that investigated speech rate entrainment during language comprehension. These studies suggest that listeners can entrain over multiple time scales (e.g., a single utterance and multiple utterances) and can predict the rate of forthcoming speech based on this entrainment (e.g., Dilley & Pitt, 2010). In other words, our separate effects of context and final word rate in Experiment 5 suggest that listeners form and sustain timing predictions over long time scales (i.e., multiple syllables), but can also adjust their predictions rapidly over shorter time scales (i.e., a single syllable). Crucially, our experiments demonstrate that entrainment over multiple timescales during comprehension can influence the timing of later production, which is consistent with a tightly yoked account. In other words, timing representations are shared across production and comprehension and listeners use speech rate entrainment during comprehension to time articulation of responses during language production, consistent with Garrod and Pickering (2015).

Our results suggest that entrainment facilitates coordination during conversational dialogue. These entrainment mechanisms may also be involved in coordinating multiple levels of representations during dialogue, even non-linguistic representations. For example, Shockley, Santana, and Fowler (2003; see also Shockley, Baker, Richardson, & Fowler, 2007) found that participants reading words out loud in synchrony tended to entrain their postural movements as well. Thus, the entrainment mechanisms used during language comprehension and production may also be implicated in coordinating behaviors across other modalities more generally, such as action and perception.

In Experiment 6, we found no evidence to suggest that response timing (based on speech rate entrainment) was affected by response preparation. Although further research is needed to confirm this finding (since our conclusions are based on a null interaction), we note that this result is consistent with Bögels and Levinson's (2017; see also Levinson & Torreira, 2015) early-planning hypothesis, which claims that listeners often prepare their verbal response independently from timing articulation (i.e., without necessarily knowing when they will have the opportunity to launch articulation). In instances where listeners do prepare early, they must hold this response in an articulatory buffer until they can launch articulation (Piai et al., 2015a). Conversely, there must also be instances where listeners know they can begin articulation, but have not yet prepared their response. In these instances, the listener most likely has to plan their response incrementally at the same time as they articulate earlier aspects of their response (e.g., Ferreira & Swets, 2002).

Although we focused on the timing of participants' responses (i.e., how quickly they responded), we also note that faster responses are not necessarily better. Interlocutors need not only ensure they produce their response quickly, but they must also do so without extensively overlapping with the previous speaker, in part because conversational overlap may  reduce intelligibility. In other words, listeners must ensure they produce their response both *quickly* and *precisely*. Indeed, in one previous study, we have considered both the timing and the precision (i.e., how closely participants respond to the end of the speaker's turn) of responses (Corps et al., 2018). Analyses of the precision of participants' responses for the current study are reported in the Appendix. However, we chose not to discuss response precision in this paper because entrainment makes predictions about response timing and not precision. In other words, if the speaker produces their turn at a fast rate, and the listener entrains to this rate, then we expect them to respond faster but not necessarily more precisely.

Even though our experiments provide evidence that listeners can use speech rate entrainment to time response articulation over both short and long timescales, we do not suggest that this is the only mechanism that listeners use to time articulation. Other studies have shown that listeners may also predict the speaker's turn-end (e.g., De Ruiter et al., 2006), react to turn-final cues (e.g., Gravano & Hirschberg, 2011), or even use a combination of these two mechanisms (Bögels & Levinson, 2017). It is likely that the listener uses whichever cues are available during dialogue, and so speech rate entrainment could work in parallel with these other mechanisms to help the listener time articulation.

In conclusion, we have shown that participants in a question-answering task use speech rate entrainment over multiple timescales (a single utterance and a single syllable) to time response articulation, suggesting that comprehension and production share timing mechanisms. In addition, we found that this entrainment mechanism did not affect the process of response preparation, and thereby argued that the processes involved in response preparation and articulation often occur independently during conversational turn-taking.

# 4. Study 3 Experiments 7-9: Prediction and integration during perceptual learning

## 4.1. Introduction

People are capable of understanding speech in a variety of different situations that alter what they hear. For example, they can comprehend speech produced by talkers at different rates (e.g., Miller & Liberman, 1979) and with different accents (e.g., Clarke & Garrett, 2004). In the case of more artificial distortion, such as time compression, comprehension is poor on initial presentation but increases with repeated exposure (e.g., Dupoux & Green, 1997). This adaptation is a form of perceptual learning – "relatively long-lasting changes to an organism's perceptual system that improve its ability to respond to its environment and are caused by its environment" (Goldstone, 1998, p. 586).

Research suggests that top-down (lexical) knowledge plays an important role in perceptual learning. For example, Norris et al. (2003) found that participants were better at learning ambiguous fricatives when they were embedded in words rather than non-words. But what aspects of top-down knowledge make it useful for learning? One possibility is that listeners use top-down knowledge to predict what they are going to hear (e.g., Altmann & Kamide, 1999) and then use this prediction to guide comprehension. If this is the case, then we would expect listeners to be better at learning to understand novel sounds when they are embedded in predictable rather than unpredictable contexts. Alternatively, it is possible that top-down knowledge makes it easier for listeners to integrate unfamiliar sounds into pre-existing representations, regardless of predictability (see Kutas et al., 2011, for a discussion on the debate about prediction versus integration).

In three experiments, we distinguish between these two possibilities by investigating perceptual learning of noise-vocoded speech. Noise-vocoding is an acoustic distortion that is created by dividing the speech stream into a number of frequency bands and then applying the amplitude envelope of each frequency range to band-limited noise, thus removing spectral information from the speech signal while still preserving temporal cues (R. V. Shannon et al.,  1995). Increasing the number of frequency bands used for vocoding increases intelligibility, such that speech vocoded with fewer than four bands is typically difficult to understand, speech vocoded with five to eight bands produces around 50% intelligibility (see Shannon et al., 2004), and speech with more than ten bands is readily intelligible.

In one of the first studies investigating perceptual learning of noise-vocoded speech, Davis et al. (2005) presented participants with sentences vocoded using six channels and asked them to report what they heard. Much like research using time-compression, the authors found that word report scores gradually increased over the course of 30 noise-vocoded sentences, starting at close to zero and reaching a maximum of around 70%. Since report scores were enhanced for words that participants had not heard in any of the previous sentences, this effect did not simply occur because participants were better at guessing sentence content or because of the benefit of repeated presentation. Instead, these results provide evidence for perceptual learning, and suggest that exposure to noise-vocoded sentences altered the processing of subsequent sentences.

In subsequent experiments, they investigated whether top-down knowledge aids learning using two training conditions: One in which participants were presented with the distorted sentence and then subsequently heard (Experiment 2) or read (Experiment 3) a clear version followed by the distorted sentence a second time (DCD condition), or one in which they heard the distorted sentence twice before hearing the clear version (DDC condition). The authors found that listeners who knew the identity of the distorted sentence prior to its second presentation (DCD condition) were able to report more words during the first presentation of subsequent vocoded sentences than participants who heard both versions of the distorted sentence before the clear version (DDC condition). In other words, listeners showed more rapid perceptual learning when they knew the identity of the distorted sentence prior to its second presentation. Similar results were reported for noise-vocoded words by Hervais-Adelman et al. (2008).

Together, these results suggest that top-down knowledge of the lexical content of distorted speech facilitates perceptual learning. These findings are consistent with a top-down component in the perceptual learning process (e.g., Norris et al., 2003), in which learning is driven by comparisons between the lexical representation of the clear stimulus and the distorted speech. Specifically, clear presentation of speech prior to distortion provides the auditory system with a target representation, which can be used to adjust incorrect representations so that they more closely match incoming speech and are thus more accurate. In other words, learning occurs when top-down knowledge is present because listeners can use this knowledge to precisely predict the form of what they are going to hear.

Sohoglu et al. (2012) found evidence consistent with this argument. In their study, they presented participants with words that were noise-vocoded using two, four, or eight channels. To manipulate prior knowledge, noise-vocoded words were preceded by the presentation of matching (text that matched the distorted word), mismatching (text that matched a different distorted word), or neutral (a string of 'x' characters) written text. Behavioral results showed that participants gave higher clarity ratings to (i) noise-vocoded words preceded by matching rather than mismatching or neutral text and (ii) words vocoded with more channels. Although Sohoglu et al. did not assess learning using word report scores, it is likely that enhanced clarity is a precursor for increased comprehension. Thus, these results are consistent with Davis et al. (2005).

In addition, concurrent MEG and EEG recordings showed reduced activity in the inferior frontal gyrus (which is associated with processing speech content; e.g., Scott & Johnsrude, 2003) when participants had prior knowledge of the distorted speech from matching text. This effect occurred before reduced activity in the superior temporal gyrus, which is associated with lower-level sensory processing. Thus, these results are consistent with a top-down process, in which higher-level lexical information modifies bottom-up processing. In particular, Sohoglu et al. interpret their findings in line with a predictive coding account (e.g., Arnal & Giraud, 2012; Arnal, Wyart, & Giraud, 2011), in which listeners use top-down knowledge to predict incoming sensory input. Listeners then compare these predictions to the actual input, and any differences yield an error signal, which is subsequently used to adjust later predictions so that they more closely match incoming input.

In a similar study, Blank and Davis (2016) found presenting matching text and increasing sensory detail in the vocoded speech (speech vocoded with twelve channels compared to four) both improved word report scores and reduced BOLD signals in the lateral temporal lobe. But in addition, these two factors interacted. When prior knowledge was uninformative (i.e., mismatching or neutral text), increasing sensory detail increased the amount of information represented in superior temporal multivoxel patterns (which measure how much information about the phonetic form of speech is contained in fMRI activation patterns). When prior knowledge was informative (i.e., matching text), however, increased sensory detail reduced multivoxel patterns. Together, these results are consistent with a predictive coding account, in which deviations from predicted input are represented as prediction errors. When sensory input matches prior knowledge (i.e., in the matching conditions), prediction errors is reduced which leads to reduced sensory activity. When there is a mismatch between prior knowledge and sensory input, prediction errors are increased and sensory activity also increases. In other words, listeners used the matching text preceding noise-vocoded speech to predict the likely sensory input.

All of the research that we have discussed so far has demonstrated that perceptual learning is enhanced in the presence of meaningful external feedback from stimulus repetition. In these cases, participants use the clear presentation of the stimulus to predict the form of subsequent distorted speech. However, it is also possible that such effects occur simply because it is easier to integrate representations of distorted speech after a clear presentation of the same stimulus. In other words, listeners do not use prediction to guide perceptual learning. Instead, faciliatory effects from written or auditory presentation of the clear stimulus prior to distortion (relative to conditions in which the clear stimulus is presented after distortion or in which the stimulus does not match the distorted text) could be attributed to increased ease of integrating the lexical representations of distorted speech into unfolding representations. For example, distorted words that match the previous clear presentation are more plausible than distorted words that do not, and research suggests that greater plausibility results in faciliatory effects, such as faster reading time (e.g., Rayner et al., 2004). Thus, these experiments do not allow us to tease apart perceptual learning effects reflecting ease of integration and prediction error, as the faciliatory effect could be attributed to either or both.

Such ease of integration may also explain the findings of studies that have demonstrated effects of semantic coherence on clarity ratings. For example, Signoret et al. (2018; see also Davis et al., 2011) presented participants with noise-vocoded sentences that were either semantically coherent, and thus listeners could use the utterance context to predict likely continuations (e.g., *Her daughter was too young for the disco*), or semantically incoherent, and did not provide information about the speaker's forthcoming words (e.g., *Her hockey was too tight to walk on cotton*). The authors found that clarity ratings were higher when these sentences were (i) preceded by matching rather than mismatching written text and (ii) semantically coherent rather than incoherent. Signoret et al. argue that these results suggest that both semantic and form-based predictions aid perceptual clarity, but they could also reflect ease of integration: Predictable words are likely more plausible continuations than less predictable words.

In sum, experiments showing faciliatory effects of feedback and semantic coherence on perceptual learning have tended to conflate manipulations of predictability and plausibility, and so it is unclear whether perceptual learning reflects prediction error or ease of integration. To discriminate between these possibilities, we conducted three experiments in which participants listened to question-answer sequences and were asked to type what they thought the answerer said. In Experiment 7 we assessed perceptual pop-out, which occurs when participants can more easily recognize words in noise-vocoded sentences because there is some constraint on their interpretation (e.g., Giraud et al., 2004). Experiment 8 tested perceptual learning effects using a manipulation similar to Davis et al.'s (2005) DCD condition, and Experiment 9 investigated the time-course of such learning effects.

In all experiments, questions were clearly spoken, while answers were noise-vocoded using six channels. To investigate prediction effects, we manipulated the predictability of these questions, so that they were either constraining and predicted a particular answer (e.g., *What colors are pandas?*; see Table 9) or unconstraining and did not predict a particular answer (e.g., *What colors should I paint the wall*?). To investigate integration effects, we manipulated the plausibility of the noise-vocoded answers, so that they were either plausible, and made complete sense as a possible answer (e.g., *Black and white*), or implausible, and made no sense (e.g., *Tom Hanks*).

If perceptual pop-out and perceptual learning effects reflect prediction error, then we expect an interaction between question predictability and answer plausibility. When the question is constraining, listeners can predict the form of a specific answer, which they can use to guide their interpretation of the distorted speech. Listeners' predictions are more likely to be accurate when the answer is plausible, and make sense as a continuation, but inaccurate when the answer is implausible. Thus, listeners are more likely to correctly report more words in the constraining plausible than the constraining implausible condition. In the unconstraining conditions, however, listeners cannot make highly specified target predictions of the likely answer, and so we expect a smaller difference in word report scores for the two plausibility conditions. But if perceptual learning effects reflect ease of integration, then we expect listeners to report more words when answers are plausible rather than implausible, regardless of whether questions are constraining or unconstraining. Finally, it is possible that both prediction error and ease of integration enhance perceptual learning, such that both question predictability and answer plausibility will additively influence word report scores.

Table 9. Example materials for the four conditions in Experiment 7-9.

| Question Predictability | Question | Answer Plausibility | Answer |
|---|---|---|---|
| Constraining | What colors are pandas? | Plausible | Black and white |
|  |  | Implausible | Tom Hanks |
| Unconstraining | What colors should I paint the wall? | Plausible | Black and white |
|  |  | Implausible | Tom Hanks |

## 4.2. Experiment 7

In Experiment 7, we tested the effects of predictability and plausibility on perceptual pop-out to determine whether prediction error, integration, or both influence perceptual learning. Participants listened to question-answer sequences, in which the question was clearly presented while the answer was noise-vocoded, and were asked to type what they thought the answerer said. Thus, participants could use the clear question to guide their interpretation of the distorted answer. Question-answer sequences fell into one of four conditions. Questions in the constraining conditions predicted a particular answer, such that the majority of participants in a pre-test converged on an answer. In the unconstraining conditions, however, questions did not predict a particular answer and participants diverged on responses in the pre-test. Importantly, answers were either plausible, and made complete sense as an answer, or implausible, and made no sense.

### 4.2.1. Method

#### *4.2.1.1. Participants*

Eighty native English speakers (21 males; *M*age = 28.56) from Prolific Academic participated in exchange for £1.70. All participants resided in the United Kingdom and had a minimum 90% satisfactory completion rate from prior assignments. Participants reported no known speaking, reading, or hearing impairments.

#### *4.2.1.2. Materials*

We selected 124 question-answer sequences (31 for each condition) using two norming tasks. First, we selected questions for the two predictability conditions using an online question-answering task, in which 31 further participants from the same population as the main experiment (8 males; *M*age = 20.67) were presented with 62 questions and were instructed to: “type your answer into the box below each question. If you do not know the answer, then please guess; do not use Google”.

Although the content predictability of utterances is typically assessed using Cloze probability (e.g., Taylor, 1953), this measure can be computed only for answers consisting of a single word because answers may differ verbatim while having the same content (e.g., *it hit an iceberg* vs *hitting an iceberg*). Answers in our task consisted of at least two words, and thus we assessed the content predictability of questions using Latent Semantic Analysis (LSA; Deerwester et al., 1990) matrix comparisons using the general reading corpus. LSA determines the semantic similarity of words and phrases by calculating the extent to which they occur in the same context, and ranges from 1 (answers are identical and the question thus constrains the answer) to -1 (answers are completely different and the question is unconstraining). Using these LSA comparisons, we calculated the predictability of each question by averaging over the LSA scores for all pairwise comparisons between answers. Questions in the constraining conditions had higher question LSA than those in the unconstraining conditions ($p < .001$; see Table 10). Note that we used the same questions in the plausible and implausible conditions, and thus Question LSA was identical across Answer Plausibility.

Table 10. Means and (standard deviations) of question LSA scores and answer plausibility for stimuli in the four conditions in Experiments 7-9.

| Question Predictability | Answer Plausibility | Question LSA[a] | Plausibility Rating[b] |
|---|---|---|---|
| Constraining | Plausible | .86 (.11) | 6.57 (0.47) |
| | Implausible | .86 (.11) | 1.31 (0.26) |
| Unconstraining | Plausible | .33 (.15) | 6.09 (0.71) |
| | Implausible | .33 (.15) | 1.68 (0.82) |

[a] Average over all answer comparisons for that particular question.

[b] Plausibility ratings made on a scale of 1-7. 1 indicated that the question was very implausible, while 7 indicated that the question was very plausible.

Using responses from the question-answering task, we selected answers (between two and four words in length) for questions in the constraining plausible conditions. Since pop-out effects may differ for different distorted stimuli (i.e., some answers may be easier to understand than others when distorted), we used the same answers in the unconstraining plausible condition, even though only 10% of these corresponded to a response that participants actually provided to the unconstraining questions in the norming task. In other words,  these answers were very rarely predicted by participants. For the two implausible conditions, we randomly rotated answers from the two plausible conditions. Thus, there were four versions of each stimulus (see Table 9).

To assess answer plausibility, we conducted a second online norming task in which 44 further participants from the same population (11 males; *M*age = 20.02) were presented with 31 question-answer sequences. We randomly assigned participants to one of four lists, created using a Latin Square procedure, so that they saw only one version of each item. Participants were instructed to: "rate the plausibility of each answer, given the preceding context of the question". Ratings were made on a scale of 1-7, where 1 indicated that the answer was very implausible (i.e., made no sense and not a possible answer to the question asked) and 7 indicated that the answer was very plausible (i.e., made complete sense and was a possible answer to the question).

Although answers in the plausible conditions had higher plausibility ratings than those in the implausible conditions ($p < .001$; see Table 10), there was also a significant interaction between question predictability and answer plausibility. In particular, answers in the constraining plausible condition had higher plausibility ratings than those in the unconstraining plausible condition ($p = .02$) and answers in the constraining implausible condition had lower plausibility than those in the unconstraining implausible condition ($p = .002$). This interaction cannot be attributed to collinearity between question LSA and plausibility ratings, since we found no correlation between these two values ($r = .013$, $p = .89$).

To try and overcome the differences in plausibility ratings in the four conditions, we conducted a second pre-test of answer plausibility, using a different set of rotated answers for the implausible conditions. However, we still found the same interaction between plausibility ratings and question LSA. Thus, it is likely that we were unable to balance plausibility ratings across the four conditions because predictability made it easier to identify implausibility. When questions are constraining, for example, there is often only one possible answer (e.g., *When is New Year's Eve? The thirty first of December*), and so all others are considered implausible because they are likely incorrect. When questions are unconstraining (e.g., *What is your favorite film?*), however, there are a variety of possible answers and so it is not particularly clear which answers are implausible. In other words, it may be easier to identify an implausible answer when the question is constraining rather than unconstraining, which leads to lower plausibility ratings for constraining implausible than unconstraining implausible questions-answer sequences. Conversely, constraining plausible question-answer sequences have higher plausibility ratings than unconstraining plausible items because it is clear what is considered a plausible answer. We return to this issue in the Data Analysis and Results sections.

Questions were recorded by a native English female speaker, who was instructed to read the utterance as though "you are asking a question and expecting a response". Answers were recorded separately by a native English male speaker, who was instructed to read the utterances as though "you are answering a question". The amount of sensory detail available in answers was varied using noise-vocoding (R. V. Shannon et al., 1995), which divides the speech signal into frequency bands and then applies the amplitude envelope in each frequency band onto corresponding frequency regions of white noise. Vocoding was performed with a custom MATLAB (MathWorks) script using six spectral channels logarithmically spaced between 70 and 5000 Hz, and thus answers were unintelligible to naïve listeners (see Davis et al., 2005).

### *4.2.1.3. Procedure*

The experiment was controlled using jsPsych (de Leeuw, 2015) and data was recorded using MySQL (version 5.7). Participants were warned that they would be listening to audio stimuli, and so were encouraged to complete the experiment in a quiet environment or to use headphones. To make stimulus onset salient, a fixation cross appeared 500 ms before question playback. The screen then turned red and answer playback began 500 ms later. Participants were told: "First you will hear a female speaker ask a question in a clear voice. You will then hear a male answer this question in a distorted voice. Your task is to listen carefully and type exactly what you think the male speaker said. If you do not know, then please guess". After typing their response, participants pressed a "submit answer" button to move onto the next trial.

### *4.2.1.4. Design*

Question predictability and answer plausibility were manipulated within items but between participants. Participants were randomly assigned to one of eight stimulus lists, each containing 15 items (one item was discarded to ensure there were an equal number of stimuli in each list), in which all items belonged to one of the four stimuli conditions. We created eight lists of 15 stimuli, rather than four lists of 31 stimuli, to ensure that answers in the implausible conditions appeared in a separate list from their corresponding question, so that they could not be primed by previous exposure (e.g., the implausible answer *James Bond* would not be primed by an earlier trial such as *Which character is also known as 007*? *King's Cross*). Participants thus heard only one version of each answer (either plausible or implausible) and one version of each question (either constraining or unconstraining), and all the items they heard belonged to the same condition. Although we assigned participants to one of four conditions, we used the continuous values of question predictability (question LSA) and answer plausibility (answer plausibility rating) when analyzing the results to overcome the differences in answer plausibility in the constraining and unconstraining conditions.

### 4.2.2. Data Analysis

For each answer, we calculated the proportion of words each participant correctly identified. Any obvious spelling mistakes or typing errors (i.e., from keys around the target letter/word, missing letters, etc.) were considered correct, but morphological mismatches were not (i.e., *younger* would be considered incorrect if the target word was *young*; see also Davis et al., 2005). Words reported in the right order were considered correct, even if intervening words were absent or incorrectly reported. Words reported in the wrong order were not scored as correct. Of the 1200 responses, we discarded 14 (1.12%) because participants typed the question rather than the distorted answer.

To evaluate the effects of question predictability and answer plausibility on the proportion of words correctly identified, we analyzed the data with generalized linear mixed effects models (GLMM; Baayen et al., 2008) using the maximal random effects structure justified by our design (Barr et al., 2013). All analyses were conducted using the *glmer* function of the *lme4* package (version 1.1-14; Bates, et al., 2015) in RStudio (version 0.99.903) using a binomial family.

For clarity, we plot the proportion of words participants correctly identified by Question Predictability and Answer Plausibility. But since there was a difference in the average plausibility of the constraining and unconstraining conditions, we did not bin items into factorial conditions when analyzing the data and instead treated Question Predictability and Answer Plausibility as continuous variables. Thus, the proportion of words correctly identified were predicted by question LSA, plausibility rating, and their interaction. Since previous research suggests that distorted speech comprehension improves over time (e.g., Davis et al., 2005), we also included Block (and its interaction with Question Predictability and Answer Plausibility) as a numeric predictor. The trials were split into three blocks of five: Block 1 included trials 1-5, Block 2 included trials 6-10, and Block 3 included trials 11-15. All predictors were centered before being added to the model.

We report the coefficient estimates (*b*), standard error (*SE*), and *p* values for each predictor. In addition, we computed the Bayes factors for all predictors by fitting generalized Bayesian mixed effects models using the *brms* package (version 2.1.0; Bürkner, 2018) with student_t priors (with ten degrees of freedom, a mean of zero, and a standard deviation of one) for all population-level effects. In all instances, we compared the full model to a model excluding the relevant predictor(s). Following Dienes (2014), we interpret a Bayes factor (i) greater than 3 as strong evidence for the alternative hypothesis over the null, (ii) less than 0.33 as strong evidence for the null hypothesis over the alternative, and (iii) between 0.33 and 3 as weak evidence.

### 4.2.3. Results

On average, participants correctly identified 60% (0.60) of the words in the distorted answers (see Fig. 14 for a breakdown of proportions by condition and block). Our analysis (see Table 11) showed that participants were better able to report the words in the answer when that answer was a more rather than less plausible response to the preceding question (effect of Answer Plausibility, $b = 2.74$, $SE = 0.33$, $p < .001$). Overall, participants were not any better at reporting the words in the answer when questions were more constraining (effect of Question Predictability, $b = 0.26$ $SE = 0.24$, $p = .28$). However, we did find an interaction between the constraint from the question and the plausibility of the answer ($b = 0.80$, $SE = 0.24$, $p < .001$), such that having a constraining question improved performance when reporting more plausible answers, but did not improve performance when answers were implausible. This interaction is illustrated in Fig. 15, which shows that there was a positive relationship between Question Predictability (Question LSA) and the proportion of words participants correctly identified in the answer at higher plausibility ratings (i.e., above 4), but a negative relationship at lower plausibility ratings (i.e., below 4).

Figure 14. Observed means of the proportion of words correctly identified for the four factorial conditions across the three blocks in Experiment 7. Error bars represent ± 1 standard error from the mean.

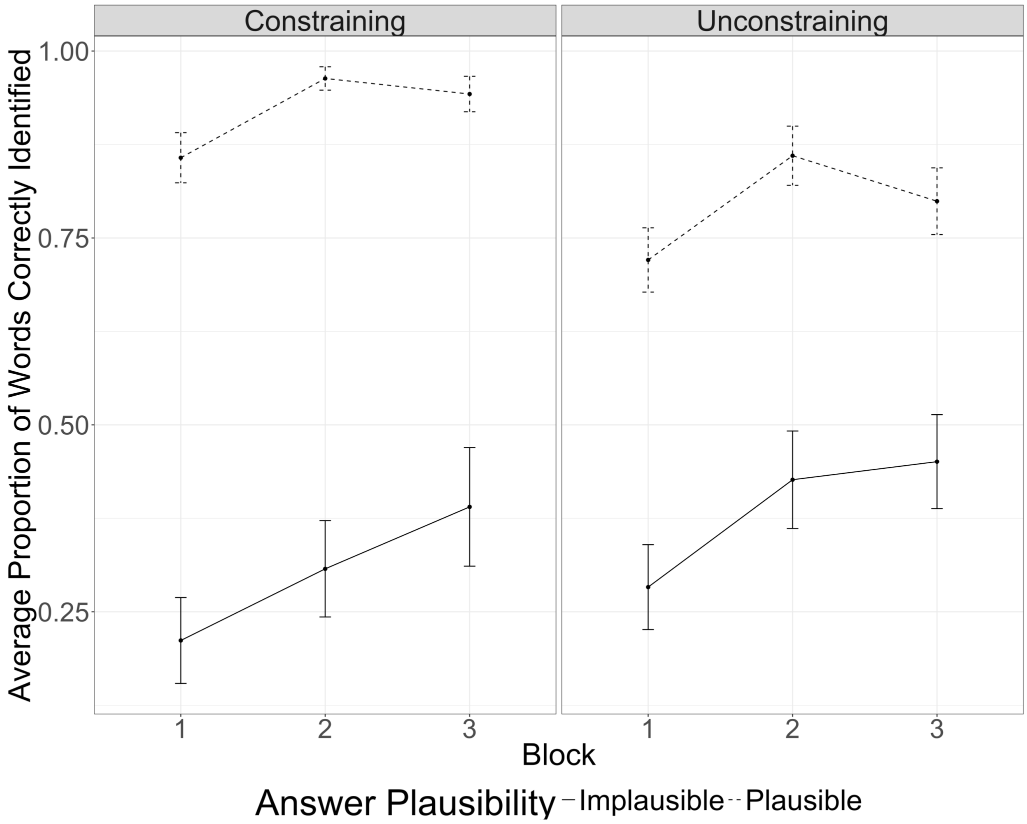
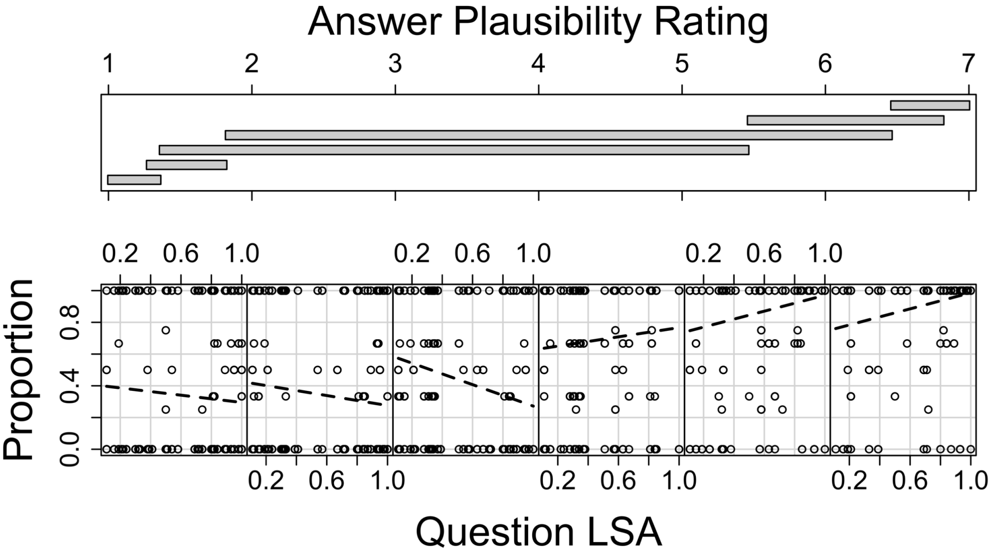


Table 11. Full model output for fixed effects for the analysis of word report scores in Experiment 7.

| Predictor | Estimate (*SE*) | *z* | *p* value | Bayes factor |
|---|---|---|---|---|
| Intercept | 1.21 (0.43) | 2.83 | .005 | - |
| Question Predictability | 0.26 (0.24) | 1.09 | .28 | 0.33 |
| Answer Plausibility | 2.74 (0.33) | 8.30 | < .001 | 5530030279 |
| Block | 0.76 (0.17) | 4.58 | < .001 | 2895 |
| Question Predictability * Answer Plausibility | 0.80 (0.24) | 3.32 | < .001 | 25.16 |
| Question Predictability * Block | 0.21 (0.19) | 1.08 | .28 | 0.43 |
| Answer Plausibility * Block | -0.01 (0.19) | -0.06 | .96 | 0.36 |
| Question Predictability * Answer Plausibility * Block | 0.12 (0.21) | 0.55 | .58 | 0.27 |

Figure 15. The relationship (represented by points and regression lines) between the proportion of words correctly identified and Question LSA at each level of Answer Plausibility Rating in Experiment 7. Note that each point represents a trial.



This interaction is consistent with a prediction error account. However, this interaction may have also occurred simply because participants were not sure what they heard the answerer say, and so they typed what they expected to hear given the context of the question. In other words, participants were worse at reporting the words in the implausible answers when questions were constraining rather than unconstraining because they were biased towards reporting an answer that made sense given the preceding question rather than the answer they actually heard.

We investigated this possibility by determining the proportion of false alarms (i.e., trials on which participants typed the plausible expected answer rather than the implausible heard answer) participants reported for the constraining and unconstraining implausible trials. False alarms occurred more often in the constraining implausible conditions (76 trials; 25%) than the unconstraining implausible conditions (eight trials; 3%) and participants reported a greater number of words in the predicted answer in the constraining implausible (*M* = .20) than the unconstraining implausible condition (*M* = 0.02; *b* = 1.63, *SE* = 0.49; *p* < .001; tested by fitting a GLMM in which the proportion of words identified in the expected answer (i.e., false alarms) was predicted by Question Predictability, with by-item slopes includes for this predictor). Thus the interaction between Question Predictability and Answer Plausibility may have occurred not because participants heard the predicted answer, but because they were biased towards reporting answers consistent with the question because they were not sure what the answerer said.

Consistent with previous research (e.g., Davis et al., 2005), we also found that participants were better at identifying the words in the later than the earlier blocks (effect of Block; *b* = 0.76, *SE* = 0.17, *p* < .001). However, Block did not interact with Question Predictability (*b* = 0.21, *SE* = 0.19, *p* = .28) or Answer Plausibility (*b* = 0.01, *SE* = 0.19, *p* = .96), and there was no three-way interaction between these predictors (*b* = 0.12, *SE* = 0.21, *p* = .58).

### 4.2.4. Discussion

In Experiment 7, we investigated whether perceptual pop-out is driven by prediction error or integration. Participants listened to question-answer sequences, in which the answer was noise-vocoded. To investigate the role of prediction, we manipulated the predictability of questions, so that they constrained a particular answer (e.g., *What colors are pandas?*) or were similarly sensible but did not constrain a particular answer (e.g., *What colors should I paint the wall?*). To investigate integration, we manipulated the plausibility of answers, so that they were either plausible and made sense as a possible response given the context of the question (e.g., *Black and white)*, or implausible and made no sense (e.g., *Tom Hanks*).

We found that participants were better at reporting words in distorted answers when they were rated as more rather than less plausible in a pre-test, regardless of question predictability, suggesting that hearing a distorted stimulus that made sense as a possible answer to a previously presented question induced perceptual pop-out. In other words, this effect is consistent with an account in which top-down information induces perceptual pop-out by increasing ease of integration.

We also found that word report scores were unaffected by question predictability (and the Bayes factor confirmed this null effect; see Table 11). One possible reason for this lack of effect is that performance in the plausible conditions was close to ceiling (see Fig. 14), thus preventing a difference between the constraining and unconstraining conditions. However, we did find that predictability enhanced perceptual pop-out at higher levels of answer plausibility. This interaction is consistent with a prediction error account, in which listeners use top-down knowledge to generate predictions about the likely form of the distorted input. However, follow-up analyses showed that this interaction likely occurred because participants were biased towards reporting answers consistent with the question in the constraining conditions, which meant that they interpreted the heard answer incorrectly when it was implausible. In other words, participants were biased towards reporting the answer they expected to follow the question rather than what they actually heard. This effect did not occur for the unconstraining conditions because the question did not place any specific constraint on the answer.  Thus, this interaction does not provide convincing support to suggest that prediction error plays a role in perceptual pop-out.

We also found that word report scores increased across the 15 trials (three blocks of five trials), suggesting that the way listeners comprehended distorted speech changed with repeated exposure. This result is consistent with previous studies using noise-vocoded speech (e.g., Davis et al., 2005) and other forms of distortion, such as time compression (e.g., Dupoux & Green, 1997). In Experiment 2, we use a training procedure similar to Davis et al. (2005) to investigate this adaptation in more detail. This design also removes the influence of response bias, thus allowing us to further investigate the interaction between question predictability and answer plausibility.

## 4.3. Experiment 8

The results of Experiment 7 suggest that perceptual pop-out is driven by answer plausibility, consistent with an integration account of perceptual learning. We also found that question predictability enhanced perceptual pop-out at higher levels of answer plausibility, but this effect reflected response bias rather than prediction error. In Experiment 8, we sought to further establish what role prediction and integration play in perceptual learning by using the design of Davis et al. (2005). In particular, we used the same stimuli as Experiment 7, but instructed participants to report the noise-vocoded answer before hearing its corresponding question. After reporting each vocoded answer, participants heard the corresponding question presented as clear speech and then the vocoded answer a second time, allowing them to use that question context to learn to process the answer. Thus, by measuring word report scores to noise-vocoded answers prior to hearing clear questions, this design assesses perceptual learning without assessing pop-out processing (and thus removes the issue of response bias).

### 4.3.1. Method

#### *4.3.1.1. Participants*

One hundred and twenty-eight further native English speakers (25 males; *M*age = 20.47) participated on the same terms as Experiment 7. We first recruited 100 participants (19 males; *M*age = 18.44) from the undergraduate student pool at the University of Edinburgh. who participated in exchange for partial course credit. Using the same procedure as Experiment 1, we recruited the remaining 28 participants (6 males; *M*age = 27.71) from Prolific Academic. We used two different participant samples because some testing occurred outside of semester time, and so we could not recruit all participants in exchange for course credit.

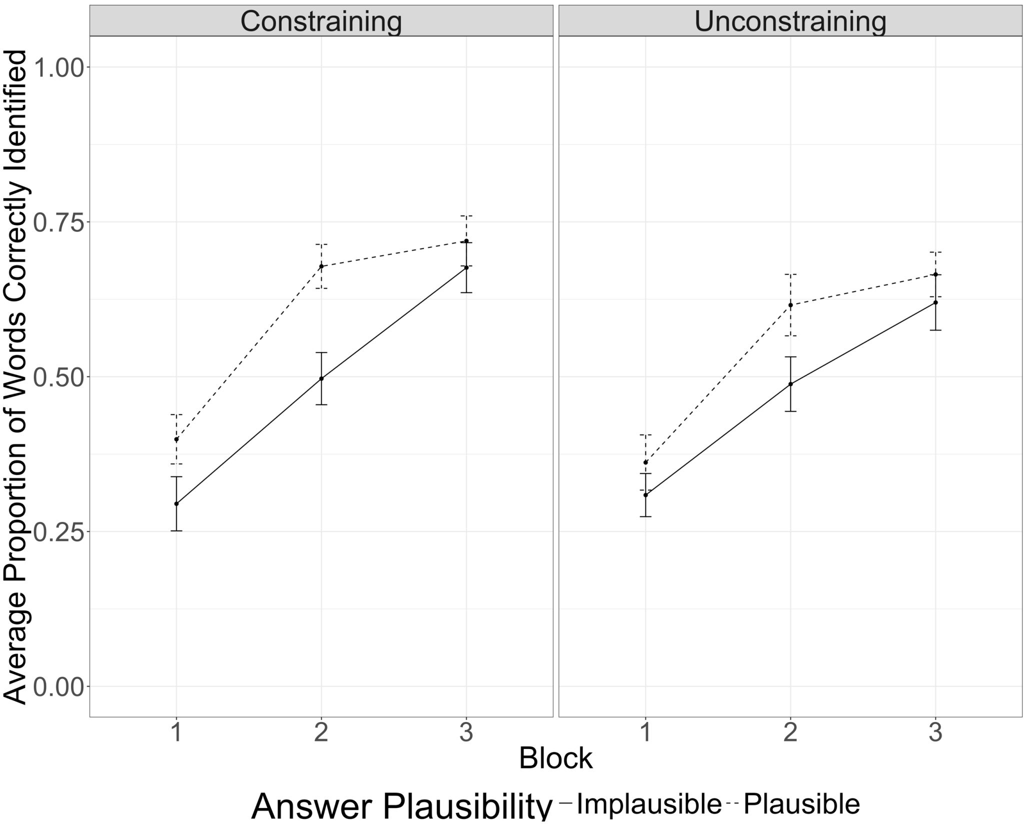
#### *4.3.1.2. Materials and Procedure*

The materials were identical to those used in Experiment 7. Participants were tested using the same procedure as Experiment 7, but they were first presented with the distorted answer, followed by the clear question, and then the same distorted answer a second time. Participants were told: “First, you will hear a male speaker produce a statement in a distorted voice. Please type the words of that statement in the box provided. You will then hear a female speaker produce the question to that statement in a clear voice. The male speaker will then repeat the distorted statement a second time. You do not need to type this statement a second time; please just listen to the exchange". To make stimulus onset salient, the screen turned red 500 ms before each answer was played for the first time. After typing their response to the first answer, participants pressed the enter key and a black fixation cross appeared 500 ms before question playback. After question playback, a red fixation cross appeared 500ms before answer onset. Participants were then prompted to press the enter key to begin the next trial.

### 4.3.2. Results

We analyzed the results using the same procedure as Experiment 7. Of the 1920 responses, we discarded six (0.31%) because participants reported the question from the previous trial rather than the answer for the current trial. On average, participants correctly identified 53% (0.53) of the words in the distorted answers (see Fig. 16 for a breakdown of proportions by condition and block).

Figure 16. Observed means of the proportion of words correctly identified for the four factorial conditions across the three blocks in Experiment 8. Error bars represent ± 1 standard error from the mean.



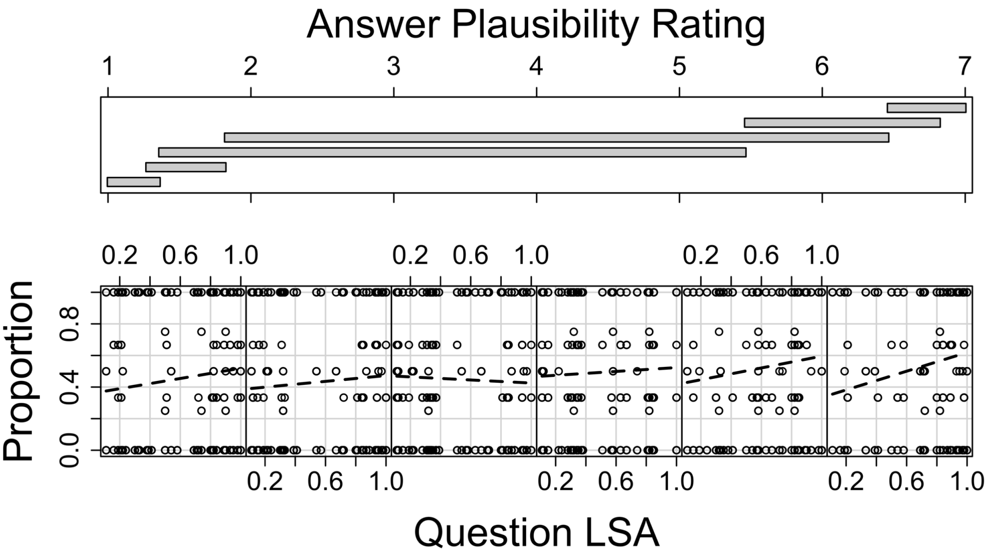
Consistent with Experiment 7, we found that participants were better at reporting words in the distorted answers when they were trained with question-answer sequences in which the answer was more rather than less plausible response to the question (effect of Answer Plausibility, *b* = 0.30, *SE* = 0.11, *p* = .004; see Table 12). Additionally, word report scores in the distorted answers did not differ for constraining and unconstraining questions (effect of Question Predictability, *b* = 0.05, *SE* = 0.10, *p* = .61).

Table 12. Full model output for fixed effects for the analysis of word report scores in Experiment 8.

| Predictor | Estimate (*SE*) | *z* | *p* value | Bayes factor |
|---|---|---|---|---|
| Intercept | 0.27 (0.28) | 0.98 | .33 | - |
| Question Predictability | 0.05 (0.10) | 0.50 | .61 | 0.18 |
| Answer Plausibility | 0.30 (0.11) | 2.89 | .004 | 8.36 |
| Block | 0.94 (0.10) | 9.24 | < .001 | 113557430 |
| Question Predictability * Answer Plausibility | 0.02 (0.10) | 0.18 | .86 | 0.15 |
| Question Predictability * Block | -0.02 (0.08) | -0.19 | .85 | 0.00 |
| Answer Plausibility * Block | 0.04 (0.08) | 0.52 | .60 | 0.19 |
| Question Predictability * Answer Plausibility * Block | -0.17 (0.09) | -1.81 | .07 | 0.98 |

Unlike Experiment 7, however, there was no interaction between the constraint from the question and the plausibility of the answer on learning (*b* = 0.02, *SE* = 0.10, *p* = .86), such that word report scores were similar for plausible and implausible answers, regardless of question predictability. This interaction is illustrated in Fig. 17, which shows that there was a positive relationship between Question Predictability (Question LSA) and the proportion of words participants correctly identified in distorted answers at all plausibility ratings, except for questions with a plausibility rating between 3 and 4.

Figure 17. The relationship (represented by points and regression lines) between the proportion of words correctly identified and Question LSA at each level of Answer Plausibility Rating in Experiment 8. Note that each point represents a trial.



Additionally, participants were better at identifying words in distorted answers in the later than the earlier blocks (effect of Block *b* = 0.94, *SE* = 0.10, *p* < .001). Although Block did not interact with Question Predictability (*b* = 0.02, *SE* = 0.08, *p* = .85) or Answer Plausibility (*b* = 0.04, *SE* = 0.08, *p* = 0.60), there was a marginally significant three-way interaction between these predictors (*b* = 0.17, *SE* = 0.09, *p* = 0.07). To follow-up this interaction, we fitted separate models for each block. We found that participants were marginally better at identifying words in distorted answers with higher Answer Plausibility in Block 2 (*b* = 0.38, *SE* = 0.21, *p* = .07) but not in Blocks 1 (*b* = 0.29, *SE* = 0.21, *p* = .15) or 3 (*b* = 0.30, *SE* = 0.23, *p* = .19). Furthermore, participants were marginally better at identifying words in distorted answers when they were preceded by questions that were more predictable in Block 3 (effect of Question LSA; *b* = 0.44, *SE* = 0.23, *p* = .06) but not in Blocks 1 (*b* = -0.02, *SE* = 0.23, *p* = .92) or 2 (*b* = 0.32, *SE* = 0.21, *p* = .13). But importantly, and inconsistent with the prediction error account, there was no interaction between Answer Plausibility and Question Predictability in any of the blocks (Block 1: *b* = -0.09, *SE* = 0.20, *p* = .66; Block 2: *b* = 0.16, *SE* = 0.22, *p* = .48; Block 3: *b* = 0.24, *SE* = 0.28, *p* = .38).[6]

### 4.3.3. Discussion

In Experiment 8, we investigated whether question predictability and answer plausibility influence perceptual learning of noise-vocoded speech. Consistent with Experiment 7, participants were better at identifying words in distorted answers when they had higher rather than lower plausibility ratings in a pre-test. However, word report scores were unaffected by question predictability and we found no interaction between question predictability and answer plausibility. Together, these results suggest that listeners were better at understanding novel distorted answers when they had been previously exposed to question-answer sequences in which the answer was a plausible rather than an implausible answer to the question, regardless of whether the question predicted a particular answer or not.

These results extend previous studies (e.g., Davis et al., 2005) and clarify how top-down knowledge aids perceptual learning. In particular, our findings are

[6] Note that we ran a version of this experiment with 64 participants (34 males; *M*age = 36.94) from Amazon Mechanical Turk. This experiment did not show any effects of Question Predictability or Answer Plausibility. However, these lack of effects likely occurred because stimuli were pre-tested on British English speakers, and many of them (e.g., *Who is the best Scottish tennis player? Andy Murray*) are culturally specific.

consistent with an integration account, in which learning occurs in the presence of informative top-down knowledge because this information makes it easier to integrate representations of distorted speech into pre-existing representations. Specifically, listeners show enhanced learning when answers were more rather than less plausibile because representations of these utterances are easier to integrate. In contrast, we did not find any effects of question predictability. This finding is inconsistent with a predictive coding account, which claims that  listeners use top-down knowledge to generate highly specified target representations (i.e., a prediction) of the distorted stimulus. Mismatches (or prediction error) between this representation and the actual stimulus are then used to adjust future predictions, so stimuli are processed more efficiently in the future. We discuss the theoretical implications of this finding in more detail in the General Discussion.

In sum, Experiment 8 demonstrates that top-down information enhances perceptual learning by increasing ease of integration rather than minimizing prediction error. In Experiment 9, we further distinguish between these two accounts by investigating the time-course of learning effects.

## 4.4. Experiment 9

Thus far, our experiments have demonstrated that top-down knowledge enhances perceptual pop-out and perceptual learning by increasing ease of integration. But in these experiments, we focused on the effect of the immediate context (from presentation of the question before the distorted answer) on perceptual learning. In Experiment 9, we further discriminate between prediction error and integration accounts by investigating the time-course of the influence of top-down knowledge on learning. To do so, we used the same procedure as Experiment 8, but presented participants with all 31 stimuli so that answers in the implausible conditions (e.g., *What colors are pandas? Tom Hanks*) could be primed by the presentation of their corresponding question many trials previously (e.g., *Who voices the character Woody in the movie Toy Story*?). In other words, we tested whether participants were better at reporting the words in the noise-vocoded implausible answers on their first presentation when the question relating to the answer had been presented earlier in the experiment.

If learning is driven by prediction error, then we do not expect implausible answers to be primed by the previous presentation of their corresponding question. In particular, this account predicts that listeners use the immediate context (i.e., the question) to predict the likely distorted answer, and any mismatches between their predictions and the actual distorted stimulus are used to adjust future predictions, meaning that distorted speech is more efficiently processed on future trials. If learning is driven by integration, however, then we expect long-term priming to enhance perceptual learning, since previous presentation of a question relevant to a later implausible answer will make this answer easier to integrate when it actually occurs.

### 4.4.1. Method

#### *4.4.1.1. Participants*

Sixty participants (12 males; *M*age = 21.95) at the University of Edinburgh participated on the same terms as Experiment 8.

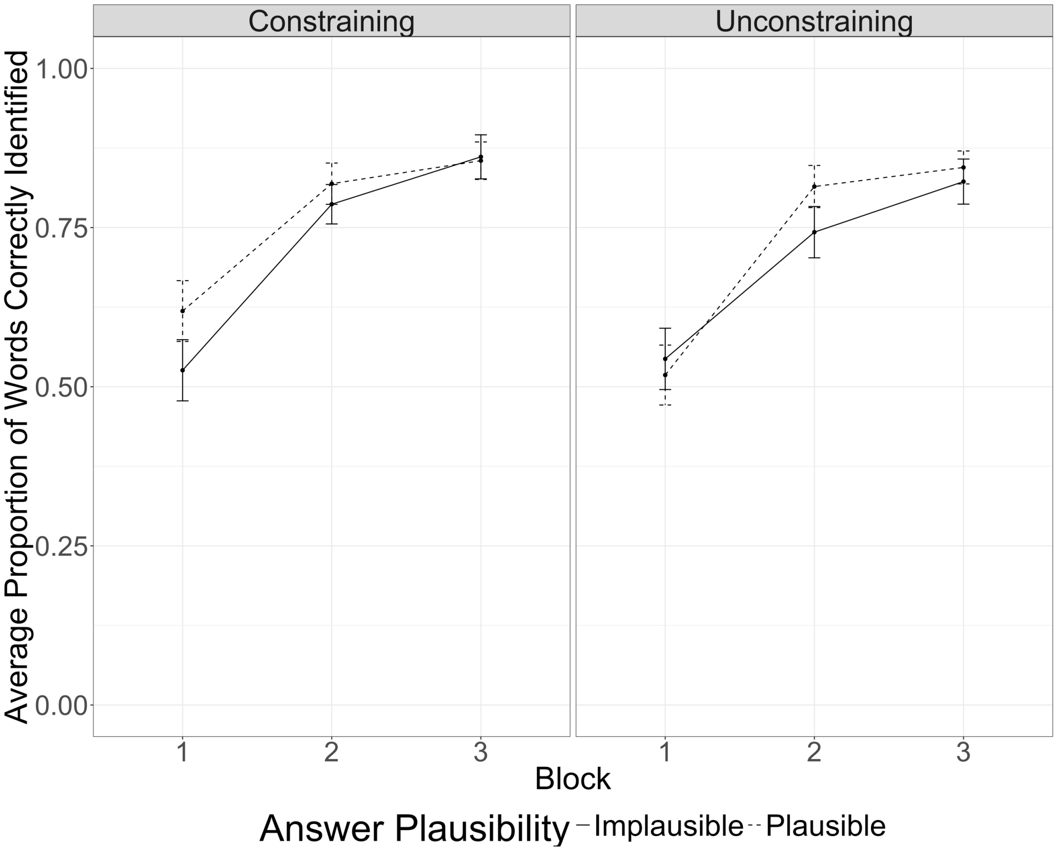
#### *4.4.1.2. Materials and Procedure*

The materials were identical to those used in the previous experiments, but we created four lists of the 31 stimuli pre-tested in Experiment 7. Thus, participants were assigned to one of four lists in which all stimuli belonged to one of the four stimulus conditions. The procedure was identical to Experiment 8, but participants were tested in-lab and the experiment was controlled using OpenSesame (version 3.0.7).

### 4.4.2. Results

The data were analyzed using the same procedure as Experiment 7, but Block 1 included trials 1-10, Block 2 included trials 11-20, and Block 3 included trials 21-31. Of the 1860 responses, eight (0.43%) were discarded because participants reported the question from the previous trial rather than the answer. On average, participants correctly identified 73% (0.73) of the words in the distorted answers (see Fig. 18 for a breakdown by condition and block).

Figure 18. Observed means of the proportion of words correctly identified for the four factorial conditions across the three blocks in Experiment 9. Error bars represent ± 1 standard error from the mean.



Unlike Experiments 7 and 8, we found that participants in Experiment 9 were not any better at reporting words in the first presentation of distorted answers when they were trained with question-answer sequences in which the answer was a plausible rather than an implausible response to the question (effect of Answer Plausibility, *b* = 0.12, *SE* = 0.25, *p* = .62; see Table 13). Additionally, participants were no better at reporting the words in the distorted answers when they were trained with constraining questions, which predicted a particular answer, than unconstraining questions, which did not (effect of Question Predictability, *b* = 0.65, *SE* = 0.38, *p* = .09).

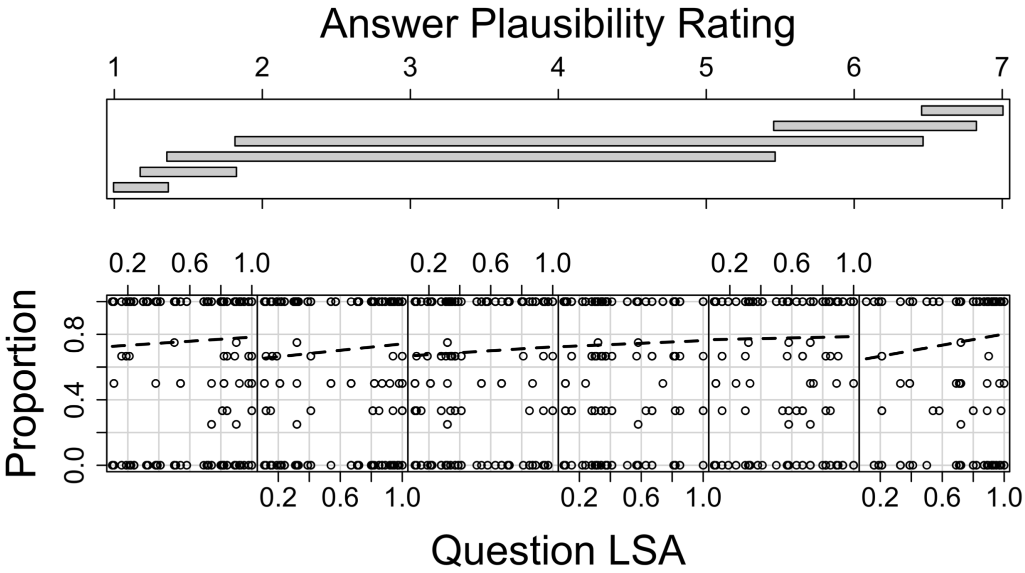
Table 13. Full model output for fixed effects for the analysis of word report scores in Experiment 9.

| Predictor | Estimate (*SE*) | *z* | *p* value | Bayes factor |
|---|---|---|---|---|
| Intercept | 1.61(0.39) | 4.12 | <.001 | - |
| Question Predictability | 0.65 (0.38) | 1.71 | .09 | 0.29 |
| Answer Plausibility | 0.12 (0.25) | 0.49 | .62 | 0.23 |
| Block | 1.03 (0.19) | 5.46 | < .001 | 107409217 |
| Question Predictability * Answer Plausibility | 0.001 (0.35) | 0.001 | .99 | 0.16 |
| Question Predictability * Block | 0.34 (0.29) | 1.17 | .24 | 0.18 |
| Answer Plausibility * Block | -0.08 (0.18) | -0.46 | .64 | 0.15 |
| Question Predictability * Answer Plausibility * Block | -0.05 (0.26) | -0.19 | .85 | 0.15 |

Consistent with Experiment 8, there was no interaction between Question Predictability and Answer Plausibility (*b* = 0.001, *SE* = 0.35, *p* = .99). Thus, word report scores were similar for plausible and implausible answers, regardless of whether the question predicted a particular answer or not (see Fig. 19). Additionally, we replicated our previous experiments and found that although participants were better at identifying words in the distorted answers in the later than the earlier blocks (effect of Block, *b* = 1.03, *SE* = 0.19, *p* < .001), Block did not interact with Answer

Plausibility (*b* = 0.08, *SE* = 0.18, *p* = .64) or Question Predictability (*b* = 0.34, *SE* = 0.29, *p* = .24), and there was no three-way interaction between these predictors (*b* = 0.05, *SE* = 0.26, *p* = .85).

Figure 19. The relationship (represented by points and regression lines) between the proportion of words correctly identified and Question LSA at each level of Answer Plausibility Rating in Experiment 9. Note that each point represents a trial.



Finally, we fitted an additional model for word report scores in the implausible conditions to test for answer priming (i.e., whether participants were better at reporting words in the distorted answers on their first presentation when the question relating to the answer had been heard earlier in the experiment). Word report scores were predicted by Question Predictability, Question Prime (reference level: after vs. before), and their interaction. We also included Block as a fixed effect to account for the possibility that any priming effect we observe may be influenced by the answer's position in the experiment, given that our previous experiments demonstrated that participants are better at reporting words in answers when they occur later rather than earlier in the experiment. Question Prime was contrast coded (-0.5, 0.5), and all predictors were centered. We again fitted models using the maximal random effects structure, which included both by-participant and by-item slopes for Question Prime and Block, and by-item slopes for Question Predictability.

Even when controlling for Block, we found that participants were better at identifying the words in an implausible distorted answer when its corresponding (i.e., plausible) question appeared before rather than after it (*b* = 0.42, *SE* = 0.21, *p* = .05), suggesting that participants' responses to distorted implausible answers were primed by the previous presentation of their corresponding question. In other words, participants activated the relevant lexical nodes necessary for interpreting the distorted answer to the question on its first presentation, which primed perception of that distorted answer when it actually occurred later on in the study.

There was no interaction between Question Prime and Question Predictability (*b* = 0.04, *SE* = 0.18, *p* = .81), suggesting that this priming effect was comparable for constraining (before *M* = 0.82; after *M* = 0.64) and unconstraining questions (before *M* = 0.78; after *M* = 0.64). Consistent with our previous analyses, participants were better at reporting words in answers that occurred in later rather than earlier blocks (*b* = 1.10, *SE* = 0.18, *p* < .001).

### 4.4.3. Discussion

In Experiment 9, we investigated the time-course of the influence of top-down knowledge on perceptual learning. To do so, we used the same procedure as Experiment 8, but presented participants with 31 stimuli, meaning that answers in the implausible conditions (e.g., *What colors are pandas? Tom Hanks*) could in principle be primed by the presentation of its corresponding question on a previous trial (e.g., *Which actor voices Woody in the movie Toy Story?*). Indeed, we found that listeners were better at reporting the words in distorted implausible answers when their question was presented multiple trials before rather than after the distorted answer. In other words, presentation of the clear question increased the degree of activation of associated lexical nodes (i.e., possible answers), which then made it easier to integrate the distorted answer when it actually occurred. We discuss the theoretical implications of this finding in more detail in the General Discussion.

We found that participants were no better at reporting the words in the plausible distorted answers compared to the implausible distorted answers. This lack of effect likely occurred because listeners could use the previous presentation of clear questions to guide their interpretation of the implausible answers, thus making it easier to integrate these representations. Finally, we replicated our previous experiments, and found that participants were no better at reporting the words in the distorted answers on their first presentation when they were trained on constraining questions, that predicted a particular answer, rather than unconstraining questions, which did not predict any particular answer.

## 4.5. General Discussion

In three experiments, we tested how top-down knowledge aids perceptual learning by presenting participants with question-answer sequences, in which the answer was noise-vocoded. We contrasted a prediction error account, in which learning is driven by a comparison process between predictions (generated on the basis of top-down knowledge) and the actual distorted input, with an integration account, which suggests that top-down knowledge of the distorted stimulus prior to its presentation makes it easier to integrate this stimulus once it is subsequently heard, thus facilitating feedback-driven learning.

We found that word report scores for noise-vocoded answers were higher when participants had been trained with question-answer sequences in which the answer was a more plausible (e.g., *Black and* white) rather than less plausible (e.g., *Tom Hanks*) a to the preceding question. Importantly, this effect occurred regardless of whether the question was constraining (e.g., *What colors are pandas*?), and listeners could use the question to predict what the answerer was likely to say, or unconstraining (e.g., *What colors should I paint the wall?*), and listeners could not predict the likely answer. We observed this effect both when participants heard the distorted answer before (i.e., perceptual learning; Experiment 8) and after (i.e., perceptual pop-out; Experiment 7) they heard the corresponding question. Finally, Experiment 9 demonstrated that the context of wider discourse could aid learning, such that participants were better able to report the words in the implausible distorted answers when their corresponding question occurred many trials before rather than after the answer.

Together, our results are consistent with previous research demonstrating that top-down knowledge, from either clear auditory or written presentation of the stimulus prior to distortion (e.g., DCD training condition in Davis et al., 2005) facilitates perceptual learning. Our results extend this work by demonstrating that top-down effects in perceptual learning reflect ease of integration, such that faciliatory effects from top-down knowledge prior to distortion occur because listeners find it easier to integrate the lexical representations of distorted speech into pre-existing representations. Our observation of improved word report scores for plausible distorted answers presented before the presentation of the clear question suggests that this adaptation does not merely occur because participants have rote learned the distorted answers or have become better at guessing their likely content. Instead, training with question-answer sequences in which the answer was plausible increased ease of integration and produced changes in pre-lexical representations, such that distorted speech was more efficiently processed in the future.

But how does ease of integration affect learning? One possibility is that upon hearing the distorted input, initial bottom-up processes activate a number of possible interpretations. Top-down knowledge, either from the presentation of a question or a clear version of the stimulus prior to distortion, then feeds back to alter pre-lexical processing to ensure that bottom-up stimulus driven processes are retuned. This retuning ensures that listeners select the most plausible interpretation and inhibit inappropriate ones, thus meaning that the perceptual system is configured to efficiently comprehend subsequently presented distorted speech. This mechanism is consistent with interactive-activation accounts of speech processing, such as TRACE (e.g., McClelland & Elman, 1986).

We observed an integration effect regardless of whether or not listeners could use the clear question to predict what the answerer was likely to say. Thus, our findings do not offer support for a predictive coding account (e.g., Sohoglu et al., 2012), which claims that listeners use top-down knowledge to generate highly specified moment-to-moment target predictions of the distorted stimulus. Under this account, listeners are then thought to use mismatches (or prediction error) between this prediction and the actual stimulus to adjust their internal representations, so that their future predictions more closely match incoming distorted speech. If prediction error was driving learning then we would have expected a larger effect of plausibility at higher levels of predictability, because listeners' predictions would be more likely to be accurate when the answer was plausible and inaccurate when it was implausible.

Experiment 9 provided further evidence that ease of integration underlies feedback-driven learning. In particular, we found that listeners were better at reporting the words on the first presentation of an implausible distorted answer when its corresponding question was presented many trials previously. This findings suggests that hearing a question relevant to a later implausible answer made it easier to comprehend this answer when it was actually presented multiple trials later, such that learning was enhanced by long-term priming. This finding is consistent with an integration account, which suggests that hearing a clear question prior to the presentation of the distorted answer should change the degree of activation of associated lexical nodes (i.e., possible answers). These lexical nodes are still active once the listener actually encounters the corresponding answer, regardless of whether it occurs immediately after or a few trials after the corresponding question, which subsequently alters feedback connections between lexical and pre-lexical representations. Prediction error accounts, in contrast, predict that learning should be restricted to the immediate context, since listeners use mismatches between their prediction (based on the question) and the distorted answer to adjust future predictions, which results in perceptual learning.

We note that although we found that question predictability enhanced perceptual pop-out when answers were more rather than less plausible continuations to the preceding question, we did not observe this effect when assessing perceptual learning in Experiment 8 or 9. This discrepancy likely occurred because participants in Experiment 7 were biased towards reporting answers consistent with the question in the constraining conditions, which meant that they interpreted the heard answer incorrectly when it was implausible and did not match question context. In the unconstraining conditions, however, a number of answers were possible and so participants were less biased towards incorrectly interpreting the implausible answers. Such an effect did not occur in Experiment 8 or 9 because word report scores were assessed on the first presentation of the distorted answer, and so participants were not biased by the constraint of the previous question.

We have demonstrated that top-down knowledge can aid perceptual learning of noise-vocoded speech by easing integration through feedback-driven learning. But are other forms of distorted speech learned in the same way? Research suggests that comprehension of time-compressed speech, a manipulation which preserves the spectral information in the signal but disrupts the temporal dimension, is poor on initial presentation but increases by up to 15% with repeated exposure (e.g., Dupoux & Green, 1997). This effect generalized to speech produced by a different talker and at a different rate, suggesting that it reflected long-term perceptual learning rather than short-term adaptation. Some studies have also observed learning effects even when listeners are trained with time-compressed sentences produced in languages they do not understand, but only if these languages share phonological features (such as syllabic structure) with the language they do speak. For example, Spanish speakers show learning effects for time-compressed sentences produced in Catalan, but not for time-compressed sentences produced in Dutch or English (e.g., Pallier, Sebastián-Gallés, Dupoux, Christophe, & Mehler, 1998; Sebastián-Gallés, Dupoux, Costa, & Mehler, 2000). Since participants had no lexical knowledge of sentences produced in an unfamiliar languages, these results suggest that learning of time-compressed speech depends on phonological rather than lexical information.

Similar findings have been demonstrated with sine-wave speech, which lacks cues necessary for grouping speech into a single auditory stream (e.g., harmonic structure and amplitude comodulation; Davis & Johnsrude, 2003). For example, Bent, Loebach, Phillips, and Pisoni (2011) showed that native-English speakers trained on German sine-wave vocoded sentences showed comparable word report scores on English sine-wave test sentences as those trained with English sine-wave sentences. This across-language transfer did not occur for participants trained with Mandarin sentences, suggesting that learning of sine-wave speech depends on phonological information. However, these studies have not used a feedback procedure to investigate the role of lexical information in learning time-compressed or sine-wave speech, and so it is possible that top-down knowledge still plays a role. In fact, even though top-down knowledge aids perceptual learning of noise-vocoded speech (e.g., Davis et al., 2005), some learning can still occur for noise-vocoded non-words (e.g., Hervais-Adelman et al., 2008).

In conclusion, our studies extend previous findings by investigating how top-down knowledge aids perceptual learning. In particular, we found that listeners were better at reporting the words in novel distorted answers when they were trained with question-answer sequences in which answers were plausible rather than implausible.

However, word report scores were not influenced by the predictability of questions. Thus, we conclude that learning occurs because top-down knowledge makes it easier for listeners to integrate representations of a distorted stimulus, thus facilitating feedback-driven learning, rather than because top-down knowledge allows listeners to make highly specified target predictions about the form of the distorted speech.

# 5. General Discussion

A number of psycholinguistic studies have shown that people predict both the content (i.e., what the speaker is likely to say; see Pickering & Garrod, 2018) and timing (i.e., the rate at which an utterance is likely to be produced; see Arnal & Giraud, 2012) of utterances during language comprehension. But what role do these predictions play during online language use? To answer this question, this thesis examined how listeners use prediction to (i) prepare and articulate their utterances during conversational turn-taking, and (ii) comprehend utterances under difficult listening conditions, such as when speech is distorted. The following sections first provide an overview of the findings from the three studies presented in Chapters 2, 3, and 4 respectively (Section 5.1) before interpreting these findings in relation to theories of conversational turn-taking and perceptual learning in more detail (Section 5.2).

## 5.1. Summary of empirical findings

### 5.1.1. The role of content and length predictions in turn-end prediction and response preparation

In the first set of Experiments (Study 1; Experiments 1-4), we used button-pressing and question-answering tasks to directly compare the mechanisms underlying turn-end prediction and response preparation. We manipulated both the content predictability (i.e., the predictability of the words of the speaker's turn) and length predictability (i.e., the predictability of the number of words the speaker will use) of simple yes/no questions. We showed that listeners responded earlier in the question-answering task when the final word(s) of the question was predictable (e.g., *Are dogs your favorite* ***animal****?*) rather than unpredictable (e.g., *Do you enjoy going to the* ***supermarket****?*). However, we found no effects of content or length predictability on the precision (i.e., how closely participants responded to the speaker's turn-end) of participants' button-presses or verbal responses.

Consistent with previous research on prediction during language comprehension (e.g., Altmann & Kamide, 1999), these experiments demonstrate that listeners can use the content of a speaker's utterance to predict how it is likely to continue. But in addition, our findings suggest that listeners use such content predictions to prepare a response early in the speaker's turn. In contrast, listeners do not appear to predict the speaker's turn-end or use this prediction to time articulation.

### 5.1.2. The role of speech rate entrainment in timing response articulation

Experiments 5 and 6 (Study 2) used a question-answering task and demonstrated effects of speech rate entrainment on the timing of articulation. In particular, we manipulated the speech rate of the context (e.g., *Do you have a…*) and the final word (e.g., *dog*?) of questions using time-compression, so that each component was spoken at the natural rate or twice as fast. We found that listeners responded earlier when the context was speeded rather than natural. In other words, they entrained to the speaker's context rate during comprehension, which in turn influenced when they launched articulation. These findings are consistent with research demonstrating speech rate priming (e.g., Jungers & Hupp, 2009), but in addition suggest that entrainment influences not only the rate of subsequent utterance production, but also when an utterance is produced.

In addition to this context effect, we also found that participants responded earlier when the speaker's final word was speeded rather than natural, regardless of context rate, which is consistent with accounts that suggest listeners adjust their entrainment after a single syllable (e.g., Giraud & Poeppel, 2012). These findings are also consistent with research demonstrating that listeners entrain over multiple time scales (e.g., a single utterance and multiple utterances; Baese-Berk et al., 2014) during language comprehension and then use this entrainment to predict the rate of forthcoming speech (e.g., Dilley & Pitt, 2010). This entrainment was unaffected by the predictability of the speaker's utterance, suggesting that response preparation and articulation timing may be two independent processes. Together, these results are consistent with accounts that suggest listeners use speech rate entrainment to time response articulation during conversational dialogue (e.g., Garrod & Pickering, 2015) and demonstrate that entrainment over multiple time scales during comprehension can immediately influence the timing of later production.

### 5.1.3. Effects of prediction and integration during perceptual learning

Experiments 7-9 (Study 3) looked at the perceptual learning of noise-vocoded speech. To do so, we presented participants with question-answer sequences, in which questions were clearly spoken while answers were noise-vocoded. We manipulated the predictability of questions, so they were either constraining and predicted a particular answer (e.g., *What colors are pandas?*) or unconstraining and did not predict a particular answer (e.g., *What colors should I paint the wall?*). Noise-vocoded answers were either plausible, and made complete sense as a possible answer (e.g., *Black and white*), or implausible and made no sense (e.g., *Tom Hanks*).

We found that word report scores for noise-vocoded answers were higher when participants were trained with question-answer sequences in which the answer was more rather than less plausible response to the preceding question. This effect occurred regardless of whether the question was constraining or unconstraining and we observed it when assessing word report scores both when distorted answers were presented after hearing the corresponding question (i.e., perceptual pop-out; Experiment 7) and when they were presented before hearing the question (i.e., perceptual learning; Experiment 8). These results are consistent with research demonstrating that greater plausibility results in faciliatory effects, such as faster reading times (e.g., Rayner et al., 2004). Finally, Experiment 9 demonstrated that the context of the wider discourse could aid learning, such that participants were better able to report the words in the implausible distorted answers when their corresponding question had been presented many trials previously.

## 5.2. General implications and future directions

### 5.2.1. Implications for models of conversational turn-taking

Most theories of conversational turn-taking agree that prediction is crucial for coordinating turns with little gap or overlap. However, these theories typically disagree on how prediction aids turn-taking. The results from Study 1 in this thesis suggest that listeners use predictions of what a speaker is going to say to prepare a response, but not to predict the speaker's turn-end. Thus, these results are consistent with the early-planning hypothesis (e.g., Barthel et al., 2016, 2017; Bögels et al., 2015) and suggest that response preparation and articulation occur independently and rely on different mechanisms. In contrast, our findings do not offer any support for the late-planning hypothesis (e.g., Sjerps & Meyer, 2015), which argues that preparation and articulation are tightly interwoven, such that listeners use predictions of what the speaker is going to say (and possibly how many words they are likely to use) to predict the speaker's turn-end, and only begin response preparation close to this moment.

Study 1 suggests that listeners do not use turn-end prediction to time response articulation. In fact, we found that responses in the button-pressing task, which has typically been used to assess turn-end prediction (e.g., De Ruiter et al., 2006), were largely driven by utterance duration, perhaps suggesting that this paradigm may not be a successful method for capturing effects of turn-end prediction that are independent of effects of duration. But in addition, we did not find any evidence that response articulation was influenced by turn-end prediction when assessing this mechanism using a verbal question-answering task, which was unconfounded by utterance duration.

Comparison of these two tasks in Study 1 highlights an important methodological point for measuring response times in future studies. In particular, we analyzed both the timing (i.e., how quickly participants responded) and the precision (i.e., how closely participants responded to the speaker's turn-end) of responses to capture two different components of the turn-taking system. In particular, analyzing response timing allowed us to capture response preparation, since participants who have prepared more of their verbal response prior to the speaker's turn-end will respond earlier than those who have prepared less of their verbal response. Response precision, instead, captures turn-end prediction, since responses closer to the end of the speaker's turn are likely to reflect more accurate predictions and thus better timing of articulation. Previous studies assessing turn-end prediction have analyzed response timing only (e.g., De Ruiter et al., 2006). However, it is not clear that an earlier response necessarily reflects better turn-end prediction. In fact, earlier responses are more likely to lead to conversational overlap than later responses, which will likely cause disruption to conversational fluency. Thus, future research assessing turn-end prediction should consider response precision.

In sum, Study 1 suggests that listeners do not use turn-end prediction to time response articulation. But they must still ensure that they articulate their pre-prepared response at the appropriate moment, so they avoid long gaps or overlaps between turns. Some research suggests that listeners launch articulation of their response reactively, after they have encountered one or more turn-final cues (e.g., falling boundary tone; see Bögels & Torreira, 2015). But importantly, these cues are not necessarily perfect predictors of a speaker switch (see Gravano & Hirschberg, 2011), and so must work in parallel with other mechanisms. In Study 2, we demonstrated that one such mechanism is speech rate entrainment based on both the rate of the context of the speaker's utterance and their final syllable. Thus, these findings suggest that timing representations are closely related during language production and comprehension, such that changes in speech rate entrainment during comprehension immediately affected subsequent production. This entrainment then helped listeners time response articulation, consistent with theories that suggest listeners use speech rate entrainment to coordinate their turns during conversational dialogue (e.g., Garrod & Pickering, 2015). Since listeners must still need to identify when the speaker will reach the end of their utterance before launching articulation of their turn based on speech rate entrainment, future research could investigate how such entrainment interacts with the presence of turn-final cues.

In both Studies 1 and 2, we demonstrated that listeners prepared their response early in the speaker's turn, before the speaker reached the end of their utterance. As a result, comprehension and production processes must overlap. Previous findings suggest that these two mechanisms share representations (e.g., Menenti et al., 2011), and so future research could investigate how listeners manage the cognitive demands of simultaneous preparation and production. Additionally, we note that Studies 1 and 2 used questions that required either a *yes* or *no* response. It is possible that listeners were sensitive to content predictability in these instances because they did not have to prepare and buffer a long response. It is possible that listeners prepare less of their response in advance when it is sufficiently complex (e.g., perhaps when it consists of multiple phrases), and so future research could investigate whether the time-course of preparation is affected by the length of the listener's utterance. These findings would be relevant to research that has demonstrated that the scope of advance planning is flexible (e.g., Konopka, 2012). However, these studies have not demonstrated that the moment *when* listeners begin preparation is also flexible. Thus, this research would shed light on how listeners manage the cognitive demands of simultaneous preparation and production.

In Studies 1 and 2, we observed response latencies much longer than the 200 ms typically reported in corpus analyses (e.g., Stivers et al., 2009). One possible reason for this discrepancy is that participants in our tasks interacted with a pre-recorded speaker. However, recent research suggests that inter-turn intervals observed in experimental settings are longer than those in natural conversations, even when participants interact with a partner in real-time (Meyer, Alday, Decuyper, & Knudsen, 2018). Thus, these experimental tasks may not capture the precision of turn-taking in natural conversations, perhaps because there are certain characteristics of natural conversation that aid response timing, and these are not present in experimental tasks. Conversely, there may be characteristics of experimental tasks that slow response timing. While naturally occurring conversation does not allow us to easily assess different theories of turn-taking (such as early-planning vs. late-planning),  future research should investigate the discrepancy between turn-taking in experimental settings and natural dialogue with a view to creating experimental tasks that can better approximate the processes involved in naturally occurring conversation.

### 5.2.2. Implications for models of perceptual learning

In Study 3 (Experiments 7-9), we investigated how top-down knowledge aids perceptual learning of distorted speech by presenting participants with question-answer sequences, in which the answer was noise-vocoded. We found that both perceptual pop-out (Experiment 7) and perceptual learning (Experiment 8) were sensitive to the plausibility of the answer, but not to the predictability of the question, which suggests that top-down knowledge aids perceptual learning by easing integration. In particular, participants showed faciliatory effects when the distorted answer was plausible rather than implausible because listeners found it easier to integrate the lexical representations of the distorted speech into pre-existing representations.

The fact that such faciliatory effects occurred not only when participants heard the distorted answers after the presentation of the clear question (i.e., perceptual pop-out), but also when they heard the distorted answers before the clear questions (i.e., perceptual learning) demonstrates that ease of integration can alter pre-lexical representations associated with speech processing. In particular, learning the mapping between question-answer sequences on previous trials alters pre-lexical representations, such that novel distorted answers are better understood on their first presentation, before the participants hears the corresponding clear question. Future research could investigate exactly how ease of integration aids perceptual learning, but our findings suggest that upon hearing the distorted input, initial bottom-up processes activate a number of possible interpretations. Top-down knowledge, either from the presentation of a question or a clear version of the stimulus prior to distortion, then feeds back to alter pre-lexical processing to ensure that bottom-up stimulus driven processes are retuned so that listeners select the most plausible interpretation when processing future instances of distorted speech. In other words, ease of integration alters feedback connections so that speech is processed more efficiently. This mechanism is consistent with interactive-activation accounts of speech processing, such as TRACE (e.g., McClelland & Elman, 1986). In contrast, our findings do not offer any support for a predictive coding account (e.g., Sohoglu et al., 2012), which claims that listeners use top-down knowledge to generate highly specified target representations (i.e., a prediction) of the distorted stimulus. Listeners then use mismatches (or prediction error signals) between this representation and the actual stimulus to adjust future predictions, so that they more closely match the incoming stimulus.

Experiment 9 demonstrated that learning was enhanced by long-term priming, thus providing further support for an integration account. Previous research has focused on whether listeners show perceptual learning when the clear version of a stimulus is presented immediately before the distorted version. Our study extends these findings by demonstrating that learning can occur even when stimuli are separated by many trials. This long-term priming effect provides further support for an integration account. In particular, hearing a clear question prior to the presentation of a distorted answer changes the degree of activation of associated lexical nodes (i.e., possible answers). These lexical nodes are still active once the listener actually encounters the corresponding answer, regardless of whether it occurs immediately after or a few trials after the corresponding question, which subsequently alters feedback connections between lexical and pre-lexical representations, so that future distorted stimuli are processed more efficiently. If learning was driven by prediction error, then we would expect it to be restricted to instances in which the question and answer are presented on the same trial because listeners use mismatches between their prediction and the distorted answer to retune future predictions, which results in perceptual learning.

In sum, our findings suggest that top-down knowledge aids perceptual learning of noise-vocoded speech by easing integration, thus facilitating feedback-driven learning. Previous research has demonstrated that listeners can also use top-down knowledge to learn to understand other forms of distorted speech, such as when comprehending sine-wave speech (e.g., Remez et al., 1981) and talkers of unfamiliar accents (e.g., Maye et al., 2008). But it is not clear whether these top-down effects reflect ease of integration or prediction error. In addition, some research suggests that other distorted speech, such as time-compressed speech, can be learned in the absence of top-down knowledge (e.g., Pallier et al., 1998). Future research could further investigate what role top-down knowledge plays in learning different forms of distorted speech and whether this learning effect reflects prediction error or ease of integration, thus providing further insight into the role prediction plays in comprehending speakers in difficult conditions.

## 5.3. Conclusion

Many psycholinguistic studies demonstrate that listeners predict a speaker's unfolding utterance during language comprehension. This thesis investigated how listeners use these predictions to coordinate their utterances during conversational dialogue. We found that listeners used content predictions (of what a speaker is likely to say) to prepare a verbal response early, before the speaker reached the end of their utterance. However, listeners did not use these content predictions to predict the speaker's turn-end, so that they could time articulation. Instead, we found that listeners timed articulation by entraining to the speaker's rate of syllable production and predicting the rate of the speaker's forthcoming syllables, suggesting that comprehension and production share timing representations. However, we did not find any evidence to suggest that listeners used predictive mechanisms to adjust their pre-lexical representations, so that they could better understand distorted speech. Instead, such learning was driven by ease integration. Together, these findings suggest that prediction plays a different role in response preparation, articulation, and comprehending utterances. In particular, our findings suggest that there is a central role for (independent) predictions of content and timing when preparing and articulating turns, but no evidence for the role of form predictions when comprehending speech.

# 6. References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, 98,* 13367-13372.

Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A, 53,* 741-764.

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,* 419-439.

Almor, A. (2008). Why does language interfere with vision-based tasks?. *Experimental Psychology, 55,* 260-268.

Altmann, G. T. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264.

Arnal, L. H. & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16,* 390-398.

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67-82.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Babiloni, C., Carducci, F., Cincotti, F., Rossini, P. M., Neuper, C., Pfurtscheler, G., & Babiloni, F. (1999). Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study. *Neuroimage, 10,* 658-665.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25,* 1546-1553.

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

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals". *Frontiers in Psychology, 8,* https://dx.doi.org/10.3389/fpsyg.2017.00393.

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology, 7,* http://dx.doi.org/10.3389/fpsyg.2016.01858.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed-effects models using "Eigein" and S4* (R package version 1.1-12). Retrieved from http://CRAN.R-project.org/package=lme4

Beattie, G. W. (1981). Interruption in conversational interaction, and its relation to the sex and status of the interactants. *Linguistics, 19,* 15-36.

Beattie, G. W., Cutler, A., & Pearson, M. (1982). Why is Mrs Thatcher interrupted so often?. *Nature, 300,* 744-747.

Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3,* 255-309.

Beňuš, S. (2009). Are we 'in sync': Turn-taking in collaborative dialogues. In *INTERSPEECH 2009 – 10th Annual Conference of the International Speech Communication Association.*

Blank, H. & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS biology, 14,* http://dx.doi.org/10.1371/journal.pbio.1002577.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition* (pp. 181-216). Orlando, FL: Academic Press.

Boersma, P. & Weenink, D. (2002). Praat: Doing phonetics by computer (version 6.0.17) [Computer software]. Retrieved June 20, 2017, from http://www.praat.org.

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders showed reduced anticipatory processing of the question. *Neuropsychologia, 109,* 295-310.

Bögels, S. & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction, 50,* 71-89.

Bögels, S., Magyari, L., & & Levinson (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5,* http://dx.doi.org/10.1038/srep12881.

Bögels, S. & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends. *Journal of Phonetics, 52,* 46-57.

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General, 143,* 295-311.

Bradlow, A. R. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106,* 707-729.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75,* B13-B25.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1482-1493.

Brown, C. M., Hagoort, P., & Ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience, 11,* 261-281.

Brown-Schmidt, S. & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review, 22,* 833-843.

Bürkner, P-C. (2017). *Brms: Bayesian Regression Models using Stan* (R package version 1.6.1). Retrieved from https://CRAN.R-project.org/package=brms

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging, 17,* 85-100.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment, 6,* 284-290.

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

Cook, A. E. & Meyer, S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 886-899.

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes, 55,* 230-240.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics, 37,* 16-28.

Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Frontiers in Psychology, 3,* https://dx.doi.org/10.3389/fpsyg.2012.00364.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19,* 141-177.

Cutler, A. & Pearson, M. (1985). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 139-155). London: Croom Helm.

Dahan, D. & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 498-513.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top-down” processes? Evidence from the time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23,* 3914-3932.

Davis, M. H. & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23,* 3423-3431.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134,* 222-241.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41,* 391-407.

Dell, G. S. & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369,* 2012394.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117-1121.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there *a* replication crisis? Perhaps. Is this an example? No: A commentary on Ito, Martin & Nieuwland (2016). *Language, Cognition and Neuroscience, 8,* 966-973.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82,* 515-535.

Dhooge, E. & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language, 66,* 163-176.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5,* https://dx.doi.org/10.3389/fpsyg.2014.00781.

Dilley, L., Morrill, T., & Banzina, E. (2013). New tests of the distal speech rate effect: examining cross-linguistic generalization. *Frontiers in Psychology, 4,* https://dx.doi.org/10.3389/fpsyg.2013.01002.

Dilley, L. C. & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664-1670.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews, 81,* 181-187.

Ding, N. & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences, 109,* 11854-11859.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85,* 761-768.

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

Duncan, S. & Niederehe, G. (1974). On signalling that it's your turn to speak. *Journal of Experimental Social Psychology, 10,* 234-247.

Dupoux, E. & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 914-927.

Engel, A. K. & Fries, P. (2010). Beta-band oscillations – Signalling the status quo? *Current opinion in neurobiology, 20,* 156-165.

Erberhard, 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,* 409-436.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41,* 469-495.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30,* 210-233.

Ferreira, F. & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46,* 57-84.

Ferreira, V. S. & Pashler, H. (2003). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1187-1199.

Finlayson, I., Lickley, R. J., & Corley, M. (2012). Convergence of speech rate: Interactive alignment beyond representation. In *Twenty-Fifth Annual CUNY Conference on Sentence Processing* (p. 24). New York, NY: CUNY Graduate School and University Center.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178-210.

Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics, 41,* 29-47.

Gambi, C., Jachmann, T. K., & Staudte, M. (2015). The role of prosody and gaze in turn-end anticipation. In D. C. Noelle, R. Dale., A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. Maglio (Eds.), *Proceedings of the annual conference of the cognitive science society* (pp. 764-769). Austin, TX: Cognitive Science Society.

Gambi, C. & Pickering, M. J. (2017). Models linking production and comprehension. In E. M. Fernández & H. Smith Cairns (Eds.), *The Handbook of Psycholinguistics.* (pp. 157-181). Oxford: John Wiley & Sons.

Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27,* 181-218.

Garrod, S. & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes, 8,* 101-126.

Garrod, S. & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53,* 181-215.

Garrod, S. & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology, 6,* http://dx.doi.org/10.3389/fpsyg.2015.00751.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2,* 103-127.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3,* https://dx.doi.org/10.3380/fpsyg.2012.00238.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., & Kleinschmidt, A. (2004). Contributions of sensory input, auditory search, and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex, 14,* 247-255.

Giraud, A. L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15,* 511-517.

Glaser, M. O. & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 875-894.

Glaser, W. R. & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118,* 13-42.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585-612.

Gravano, A. & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language, 25,* 601-634.

Griffin, Z. M. & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274-279.

Grisoni, L., Miller, T. M., & Pullermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience, 37,* 4848-4858.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics, 21,* 501-529.

Heldner, M. & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics, 38,* 555-568.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 460-474.

Howes, C., Purver, M., Healey, P. G., Mills, G., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound attributions. *Dialogue & Discourse, 2,* 279-311.

Indefrey, P. & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92,* 101-144.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86,* 157-171.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience,* 1-12.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). Why the a/an prediction effect may be hard to replicate: a rebuttal to DeLong, Urbach, and Kutas (2017). *Language, Cognition and Neuroscience,* 974-983.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 240-247.

Jongman, S. R. & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming?. *Acta Psychologica, 181,* 40-50.

Jungers, M. K. & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes, 24,* 611-624.

Jungers, M. K., Palmer, C., & Speer, S. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing, 1,* 21-35.

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,* 133-156.

Keitel, A., Prinz, W., Friederici, A. D., von Hofsten, C., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children’s development. *Journal of Experimental and Child Psychology, 116,* 264-277.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26,* 22-63.

Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology, 245,* 15-26.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation, 54,* 59-101.

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language, 66,* 143-162.

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2017). Neural entrainment determines the words we hear. *BioRxiv,* http://dx.doi.org/10.1101/175000.

Krause, J. C. & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America, 115,* 362-378.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In *Predictions in the Brain: Using Our Past to Generate a Future* (pp. 190-207).

Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology, 6,* http://dx.doi.org/10.3389/fpsyg.2015.00495.

Landis, J. R. & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33,* 363-374.

Large, E. W. & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review,* 106, 119-159.

Laszlo, S. & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61,* 326-338.

Leach, L. & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55,* 306-353.

Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics, 3*, 543-550.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38.

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends in Cognitive Sciences, 20,* 6-14.

Levinson, S. C. & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6,* http://dx.doi.org/10.3389/fpsyg.2015.0073.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. G. (2004). Clear speech perception in acoustic and electric hearing. *The Journal of the Acoustical Society of America, 116,* 2374-2383.

Local, J. K., Kelly, J., & Wells, W. H. (1986). Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics, 22,* 411-437.

Local, J. K. & Walker, G. (2012). How phonetic features project more talk. *Journal of International Phonetic Association, 42,* 255-280.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 224-234.

Luo, H. & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54,* 1001-1010.

Mädebach, A., Oppermann, F., Hantsch, A., Curda, C., & Jescheniak, J. D. (2011). Is there semantic interference in delayed naming?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 522-538.

Magyari, L., Bastiaansen, M. C. M., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind speed of response in conversation. *Journal of Cognitive Neuroscience, 26,* 2530-2539.

Magyari, L. & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology, 3,* http://dx.doi.org/10.3389/fpsyg.2012.00376.

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology, 8,* http://dx.doi.org/10.3389/fpsyg.2017.00211.

Matuschek, H., Kliegel, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type 1 error and power in linear mixed models. *Journal of Memory and Language, 94,* 305-315.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32,* 543-562.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30,* 1113-1126.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 22,* 1173-1182.

Mesgarani, N. & Chang, E. F. (2012) Selective cortical representations of attended speaker in multi-talker speech perception. *Nature, 485,* 233-236.

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology, 9,* http://dx.doi.org/10.3389/fpsyg.2018.00525.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ:Eribaum.

Miller, J. L. & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 369-378.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41,* 215-225.

Miller, J. L. & Liberman, A. M. (1979). Some effects of later-occuring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457-465.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5-6), 453-467.

Näätänen, R. (1971). Non-aging fore-periods and simple reaction time. *Acta Psychologica, 35,* 316-327.

Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology, 40,* 640-647.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife, 7,* https://dx.doi.org/10.7554/eLife.33469.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is nevery necessary. *Behavioral and Brain Sciences, 23,* 299-325.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204-238.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8,* https://dx.doi.org/10.1186/1471-2202-8-89.

Otten, M. & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming?. *Discourse Processes, 45,* 464-496.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25,* 1649-1653.

Pecenka, N. & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research, 211,* 505-515.

Peelle, J. E. & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3,* https://dx.doi.org/10.3389/fpsyg.2012.00320.

Piai, V., Roelofs, A., Rommers, J., Dahlslätt, K., & Maris, E. (2015a). Withholding planned speech is reflected in synchronized beta-band oscillations. *Frontiers in Human Neuroscience, 9,* http://dx.doi.org/10.3389/fnhum.2015.00549.

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015b). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping, 36,* 2767-2780.

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language, 64,* 404-423.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 147-165.

Pickering, M. J. & Gambi, C. (in press). Predicting while comprehending language: A theory and review. *Psychological Bulletin,* https://dx.doi.org/10.1037/bul0000158.

Pickering, M. J. & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 329-347.

Pinder, J. E., Wiener, J. G., & Smith, M. H. (1978). The Weibull distribution: a new method for summarizing survivorship data. *Ecology, 59,* 175-179.

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context rate. *Attention, Perception, & Psychophysics, 78,* 334-345.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7,* 45-56.

Power, R. J. & Dal Martello, M. F. (1986). Some criticisms of Sacks, Schegloff, and Jefferson on turn taking. *Semiotica, 48,* 29-40.

Rajah, M. N., Languay, R., & Grady, C. L. (2011). Age-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. *The Journal of Neuroscience, 31,* 17941-17954.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K., M., Head, D., Williamson, A. . . . & Acker, J. D. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral Cortex, 15,* 1676-1689.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech, 54,* 147-165.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947-949.

Riest, C., Jorschick, A. B., & De Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology, 6,* http://dx.doi.org/10.3389/fpsyg.2015.00089.

Roelofs, A. (2008). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1580-1598.

Roelofs, A. & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2,* http://dx.doi.org/10.3389/fpsyg.2011.00307.

Sacks, H., Schegloff, E. A., Jefferson, G. (1974). A simplest systematics for the organization of turn-takin for conversation. *Language, 50,* 696-735.

Sanders, A. F. (1966). Expectancy: application and measurement. *Acta Psychologica, 25,* 293-313.

Schegloff, E. A. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52-133). Cambridge: Cambridge University Press.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29,* 86-102.

Schultz, B. G., O'Brien, I., Phillips, N., & McFarland, D. H. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics, 37,* 1201-1220.

Scott, S. K. & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26,* 100-107.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension – An fMRI study. *Cerebral Cortex, 22,* 1662-1670.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 76,* 379-423.

Shannon, R. V., Fu, Q. J., & Galvin III, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol, 124,* 50-54.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303-304.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 201-208.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 326-332.

Signoret, C., Johnsrude, I., Classon, E., & Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Journal of Experimental Psychology: Human Perception and Performance, 44,* 277-285.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences, 111,* E4687-E4696.

Sjerps, M. J. & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition, 136,* 304-324.

Sohoglu, E. & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences, 113,* E1747-E1756.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience, 32,* 8443-8453.

Staub, A. & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: Evidence from either… or. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 425-436.

Staub, A., Grant., M., Astheimer, L., & Cohena, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82,* 1-17.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106,* 10587-10592.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research, 11,* 139-169.

Strömbergsson, S., Hjalmarsson, A., Edlund, J., & House, D. (2013). Timing responses to questions in dialogue. In F. Bimbot. *INTERSPEECH 2013.* Paper presented at 14th Annual Conference of the International Speech Communication Association (pp. 2584-2588). Lyon: International Speech and Communication Association.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes, 50,* 23-51.

Tauroza, S. & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11,* 90-105.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin, 30,* 415-433.

Ten Bosch, L., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication, 47,* 80-86.

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology, 6,* https://dx.doi.org/10.3389/fpsyg.2015.00284.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary related duration patterns in English. *Journal of Phonetics, 28,* 397-440.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443-467.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394-417.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 423-440.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropscyhologia, 41,* 989-994.

Wesselmeier, H., Jansen, S., & Müller, H. M. (2014). Influences of semantic and syntactic incongruence on readiness potential in turn-end anticipation. *Frontiers in Human Neuroscience, 8,* https://dx.doi.org/10.3389/fnhum.2014.00296.

Wheeldon, L. & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37,* 356-381.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectations and agreement in Spanish spoken sentences. *Neuroscience letters, 346,* 165-168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39,* 483-508.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16,* 1272-1288.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7,* 701-702.

Wilson, M. & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12,* 957-968.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M. … & Schroder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron, 77,* 980-991.

# 7. Appendix A: Experimental materials used in Experiments 1-9

## 7.1. Experimental materials used in Study 1

Table A1: List of stimuli used in Experiments 1 and 2. Completions chosen from the pre-test are *italicized.*

| Content Predictability | Length Predictability | Stimulus |
|---|---|---|
| Predictable | Single | Have you passed your driving *test*? |
| | | Do you celebrate Christmas on the twenty fifth of *December*? |
| | | Can most fish breathe under *water*? |
| | | To cook a cake, will I need to put it in the *oven*? |
| | | Is red your favourite colour? |
| | | If I wear sunglasses, will they keep the sun out of my *eyes*? |
| | | Do dogs have four legs? |
| | | Have you ever forgotten your keys and been locked out of the *house*? |
| | | Are pandas the colours black and *white*? |
| | | Have you ever seen a spider with less than eight *legs*? |
| | | Is David Cameron the prime *minister*? |
| | | At University, are you a psychology *student*? |
| | | Do you regularly borrow books from the *library*? |

| | | |
|---|---|---|
| | | Is a piano a musical *instrument*? |
| | | Should I go to the zoo if I want to see a lot of different *animals*? |
| | | Is a baby kangaroo called a *joey*? |
| | | Do you think surfers are scared of being bitten by a *shark*? |
| | | Do you think most students will pass their *exams*? |
| | | Is a Dalmatian dog black and *white*? |
| | | While eating, have you ever accidentally bitten your *tongue*? |
| | | To pay for your tuition fees, did you have to take out a student *loan*? |
| | | Are dogs your favourite *animal*? |
| | | Is Andy Murray a tennis *player*? |
| | | Either at university or school, have you ever failed an *exam*? |
| | | Should I buy my friend a present for her *birthday*? |
| | | Did you wake up before 9 o’clock this *morning*? |
| | | To keep the sun out of my eyes, should I wear *sunglasses*? |
| | | Is spring your favourite season of the *year*? |
| Predictable | Varied | If my feet are cold, should I put on *some socks?* |
| | | To pay for your studies, did you take *out a loan*? |
| | | Have you ever forgotten about an assignment and handed it in *having done it on the way to class?* |

Did The Titanic sink after *it hit an iceberg*?

Have you ever taken the blame even though you weren't *at fault*?

When eating, do you cut your food with *a knife and fork*?

Do you see your parents *at the weekend*?

When you go to restaurants, do you leave *a ten percent tip*?

To communicate with others, do deaf people have to *watch and lip read*?

Is summer your favourite *season of the year?*

Do people become werewolves when *they see a full moon*?

I don't have a watch, so could you *tell me the time please?*

Have you ever been to a casino and lost *a lot of money*?

Have you ever broken your leg and used *a crutch*?

There are no clean plates left, so could you *wash some up*?

When it is cold outside, should I wear a scarf to keep *myself warm*?

Does the dentist always tell you to brush *your teeth more*?

Should I make an optician's appointment if I think I need *new glasses?*

|  |  |  |
|---|---|---|
|  |  | As well as being a student, do you also have a *part time job*? |
|  |  | This coffee is too hot, so before I drink it should I let it *cool down a little*? |
|  |  | In your tea, would you like *milk and sugar*? |
|  |  | There’s a hole in my sock, so could you get *me new ones*? |
|  |  | The dishes need cleaning, so could you *help me clean them*? |
|  |  | I’m struggling to see, so should I get *a pair of glasses*? |
|  |  | During the night, have you ever woken up after *a nightmare*? |
|  |  | I’m going to cut my hair myself, so can you get me *a pair of scissors*? |
|  |  | In the past, have you ever been late when *you had an appointment*? |
|  |  | After an argument, have you ever slammed *a door shut*? |
|  |  | My toaster is broken, so could you *fix it please*? |
| Unpredictable | Single | Have you ever visited the city of *Paris*? |
|  |  | Are you in your third year of *marriage*? |
|  |  | Are there a lot of females in your *apartment*? |
|  |  | Do you enjoy going to the *supermarket*? |
|  |  | Today, do you think I should wear a *tie*? |
|  |  | Do you need to go to the supermarket to buy some *crisps*? |

| | | |
|---|---|---|
| | | In the past, have you had a lot of different *cars*? |
| | | Would you like to see a picture of my *spider*? |
| | | Have you ever injured your *eye*? |
| | | Have you ever seen a wild *bear*? |
| | | Do you like to eat a lot of *crisps*? |
| | | During the summer, do you like spending time at the *library*? |
| | | Do you live far away from the *beach*? |
| | | Are you really looking forward to *tonight*? |
| | | Would you like to take an evening *class*? |
| | | Is an orange the same colour as a *tiger*? |
| | | If you could get a pet, would you like to get a *tortoise*? |
| | | Should I buy a new suit for my *dance*? |
| | | Do you have any lectures on *mathematics?* |
| | | Are you very scared of *ghosts*? |
| | | Do you think you are good at *singing*? |
| | | Do most people have two *siblings*? |
| | | Do you have a big *house*? |
| | | Have you ever watched a game of *cricket*? |
| | | Have you ever been on a *plane*? |
| | | Would you like to go for a walk in the *forest*? |
| | | Have you ever played a game of *poker*? |
| | | Have you ever broken your *phone*? |
| | | Are you doing anything *important*? |
| Unpredictable | Varied | Are a lot of your friends *in the same classes*? |

Do you spend a lot of your time *with friends*?

Is your favourite book *the Hunger Games*?

Did you do anything you enjoyed and *didn't expect to*?

IS your favourite film called *The Imitation Game*?

If I want to stay warm during the winter, should I put on *multiple layers*?

Do most students finish their *studies after four years*?

Have you ever been to London to visit *the Imperial War Museum?*

In a few years, would you like to move to *the mountains*?

Is your favourite TV show *The Great British Bake Off*?

Have you ever been to the cinema to watch *the Lion King*?

Are you going to celebrate *New Year in Edinburgh*?

During your lunch break, would you like to *grab a bite to eat*?

Have you ever read a book by *Suzanne Collins*?

Have you ever read a book called *Blood Diamond*?

Do you have a lot of *free time*?

Tomorrow morning, would you like to eat *your breakfast in bed*?

Should I call the police if there is someone *acting suspiciously*?

| | | |
|---|---|---|
| | | During the evening, do you *eat dinner*? |
| | | When studying, do you like to work *in the library*? |
| | | Next week, would you like to have dinner at *that new restaurant*? |
| | | In your opinion, do you think you are *a nice person*? |
| | | Tomorrow afternoon, would you like to *play football*? |
| | | When it’s raining, should I take an umbrella to *keep myself dry*? |
| | | Have you ever been to *the zoo*? |
| | | At University, are you in *lectures a lot*? |
| | | Would you like to have a *glass of wine*? |
| | | In your spare time, have you ever listened to *heavy metal*? |
| | | In the past, have you ever tried *to ice skate*? |

Table A2: List of stimuli used in Experiments 3 and 4. Completions chosen from the pre-test are *italicized.*

| Content Predictability | Length Predictability | Stimulus |
|---|---|---|
| Predictable | Single | Have you passed your driving *test*? |
| | | Can most fish breathe under *water*? |
| | | Have you ever read a Shakespeare *play*? |
| | | Is red your favourite *colour*? |

Have you ever forgotten your keys and been locked out of the *house*?

Have you ever seen a spider with less than eight *legs*?

At University, are you a psychology *student*?

Do you regularly borrow books from the *library*?

Should I go to the zoo if I want to see a lot of different *animals*?

Do you think surfers are scared of being bitten by a *shark*?

Do you think most students will pass their *exams*?

Is a Dalmatian dog black and *white*?

When meeting someone new, do you shake their *hand*?

To pay for your studies, did you take out a *loan*?

Are dogs your favourite *animal*?

Either at university or school, have you ever failed an *exam*?

Did you wake up before 9 o'clock this *morning*?

To keep the sun out of my eyes, should I wear *sunglasses*?

Is spring your favourite season of the *year*?

Do genies grant *wishes*?

Does the Queen live in Buckingham *Palace*?

Have you ever dyed your *hair*?

Do you enjoy watching horror *movies*?

|  |  |  |
|---|---|---|
|  |  | To grow, do plants need *water*? |
|  |  | Can you type without looking at the *keyboard*? |
|  |  | Is a unicorn a horse with a *horn*? |
|  |  | Do you wash your hair every *day*? |
|  |  | To pay for your tuition fees, did you have to take out a student *loan*? |
| Unpredictable | Single | Have you ever visited the city of *Paris*? |
|  |  | Today, do you think I should wear a *tie*? |
|  |  | Do you need to go to the supermarket to buy some *crisps*? |
|  |  | In the past, have you had a lot of different *cars*? |
|  |  | Would you like to see a picture of my *spider*? |
|  |  | Have you ever injured your *eye*? |
|  |  | Have you ever seen a wild *bear*? |
|  |  | During the summer, do you like spending time at the *library*? |
|  |  | Do you live far away from the *beach*? |
|  |  | If you could get a pet, would you like to get a *tortoise*? |
|  |  | Should I buy a new suit for my *dance*? |
|  |  | Do you live in a house with other *animals*? |
|  |  | Are you happy with your *grades*? |
|  |  | Do most people have two *siblings*? |
|  |  | Have you got a big *house*? |
|  |  | Would you like to go for a walk in the *forest*? |
|  |  | Have you ever played a game of *poker*? |

| | | |
|---|---|---|
| | | Have you ever broken your *phone*? |
| | | Do you participate in a lot of *experiments*? |
| | | Do you have two *homes*? |
| | | Have you ever had to visit the hospital after injuring your *body*? |
| | | Are you allergic to *fish*? |
| | | In your opinion, do you think you are a good *cook*? |
| | | Is chocolate your favourite *treat*? |
| | | Are you in a *society*? |
| Unpredictable | Varied | When you’re studying, do you like to work *silently?* |
| | | Should I call the police if there is someone *suspicious*? |
| | | In a few years, would you like to move to *Japan*? |
| | | Do you spend a lot of your time *revising*? |
| | | Before starting your studies at University, did you take a *loan*? |
| | | When it’s raining, should I take an umbrella to *university*? |
| | | Is your favourite book *religious*? |
| | | Did you do anything you enjoyed *today*? |
| | | Have you ever read a book called *Twilight*? |
| | | Have you ever read a book by *candlelight*? |
| | | Have you ever been to the cinema to watch *Wolverine*? |
| | | Have you ever been to London to visit *family*? |
| | | Next week, would you like to have dinner at *six*? |

Have you ever been to *Greece*?

At university, are you in *psychology*?

Tomorrow morning, would you like to eat *earlier*?

In your spare time, have you ever listened to *lectures*?

In the past, have you ever tried *octopus*?

Is your favourite film *recent*?

During the evening, do you *relax*?

Would you like to learn *Mandarin*?

Would you like to climb *rocks*?

Can you play *solitaire*?

Do you get nervous when speaking *publicly*?

Have you ever been admitted to hospital to have *surgery*?

Would you like to have a *snack*?

Have you ever taken the blame even though you weren’t *responsible*?

After an argument, have you ever slammed a door *shut*?

## 7.2. Experimental materials used in Study 2

Table A3: List of stimuli used in Experiments 5. Completions chosen from the pre-test are *italicized.*

| Stimulus |
|---|
| Do most people have two *jobs*? |
| Are you happy when the weather is *dull*? |
| Have you ever been bitten by a *cat*? |
| Do you drink a lot of *juice*? |
| Would you like to go for a walk in the *rain*? |
| Do you like studying in the *dark*? |
| Do cats have two *heads* |
| Do you find lectures very *dull*? |
| During the summer, do you like spending time at the *house*? |
| Have you ever visited the city of *Rome*? |
| In your opinion, are you bad at *golf*? |
| Are you allergic to *air*? |
| Do you have a sore *thumb*? |
| Have you ever been camping in the *rain*? |
| Do you like spending time at the *bar*? |
| Is chocolate your favorite *thing*? |
| Do kangaroos have two *ears*? |
| Do you have a good relationship with your *mum*? |
| Have you ever flown a *drone*? |

Would you like to live in a different *home*?

Would you like to learn a new *phrase*?

Have you ever played a game of *cards*?

Do you have small *teeth*?

Do you need to go to the supermarket to buy some *wine*?

Do you sleep before *two*?

Do you have a big *heart*?

Do you have a pet *horse*?

Is an apple the same colour as a *rose*?

Do you often feel *stressed*?

Do you ever go to the *pub*?

At the weekend, did you do something nice for your *aunt*?

Do you have a lot of *cash*?

Do you often skip *meals*?

Would you like to go running in the *rain*?

Are there a lot of females in your *job*?

Do babies often cry when they are *young*?

Do you have four *phones*?

Have you seen my new *cat*?

Have you ever squashed a spider with a *map*?

Have you been sightseeing in *Skye*?

Is your handwriting *bad*?

Have you ever drawn a picture of a *whale*?

Have you ever had to apologise to your *boss*?

Is a pear the same colour as a *grape*?

Can you play a game of *cards*?

Do you have any *pets*?

Have you ever missed a *date*?

This morning, did you eat *eggs*?

Did you watch the tennis at *noon*?

Have you ever hurt yourself on a *plane*?

Would you like to get a new *ship*?

Do you need to buy some *shoes*?

Do you think you are good at *maths*?

Do you feel *cold*?

Have you ever failed an exam in *maths*?

Have you ever dyed your hair *blonde*?

Have you ever listened to music at a *rave*?

Would you like to take an evening *class*?

Do you ever worry about being *sick*?

Do you like to eat a lot of *sweets*?

Did you pay for your own *car*?

Should I buy a new suit for my *ball*?

For Christmas dinner, do you eat *ham*?

Have you ever been on a *date*?

Do you know how to cook *well*?

Would you like to get a *bird*?

Would you like another *car*?

Have you ever had an argument with your *dad*?

Do you have more than three *cats*?

Do you live far away from the *sea*?

Do you have high *heels*?

Can I give you a *book*?

Should I buy a nice new dress for my *ball*?

Do you spend a lot of money on *books*?

Tonight, can we stay out until *two*?

This morning, did you wake up at *noon*?

When travelling, do you get *lost*?

At University, do you study *maths*?

Can you buy me a *car*?

Do you often walk to *town*?

Do you watch a lot of *sport*?

Do you like my *car*?

Are you a big fan of *cheese*?

Do you think being a vegetarian is *cool*?

Are you free to go to the *beach*?

Would you like to have an afternoon *nap*?

Do you wear a *kilt*?

Is your favorite food *thai*?

Are your parents *nice*?

Have you ever had a bad *grade*?

In your opinion, do you think you are a good *friend*?

Are you shorter than your *dad*?

Is your hair very *fine*?

Is a grape different from a *plum*?

Have you ever won a game of *pool*?

Do you believe in *love*?

Would you like to see a picture of my *niece*?

In the past, have you had a lot of different *jobs*?

Do you have two *kids*?

Have you ever watched a game of *chess*?

Do you enjoy going to the *park*?

Have you ever seen a big *bird*?

Have you ever seen a wild *swan*?

Would you like to make an appointment with the *nurse*?

Do you own a *boat*?

Today, do you think I should wear a *tie*?

Tomorrow, would you like to wear a *watch*?

Would you like to go on holiday to *Greece*?

Are you very fond of *wine*?

When travelling, have you ever been on a *horse*?

Would you like to start attending classes on *time*?

Would you like to paint your *fence*?

Do you have a *dog*?

Do you have poor *health*?

Is your job *tough*?

Do tigers have big *heads*?

Would you like to travel to *Spain*?

Do you need a new *car*?

Do you want to buy a new *horse*?

For your age, are you *wise*?

Do you think you are *bad*?

Is an orange the same colour as a *peach*?

Do you think exercising is *fun*?

Have you had a long *trip*?

Table A4: List of stimuli used in Experiments 6. Completions chosen from the pre-test are *italicized.*

| Content Predictability | Syllable length | Stimulus |
|---|---|---|
| Predictable | 1 | Do chickens lay *eggs*? |
| | | Do dogs have four *legs*? |
| | | Is Paris the capital of *France*? |
| | | Does the president of America live in The White *House*? |
| | | Is 007 also known as James *Bond*? |
| | | Is the statue of liberty in New *York*? |
| | | Does the dentist tell you to brush your teeth twice a *day*? |

Do you wash your hair every *day*?

Have you ever seen a spider with less than eight *legs*?

Are pandas the colours black and *white*?

To pay for your studies, did you take out a *loan*?

Is a unicorn a horse with a *horn*?

Is a banana a *fruit*?

Is platform nine and three quarters at Kings *Cross*?

2 Is Harry Potter's best friend called Ron *Weasley*?

Is red your favorite *colour*?

Do genies grant *wishes*?

Did the titanic sink after hitting an *iceberg*?

Does the Queen live in Buckingham *Palace*?

Is Andy Murray a Scottish tennis *player*?

Do you think most students will pass their *exams*?

Will I need to buy a stamp before posting a *letter*?

Does the River Thames run through *London*?

Have you ever lived in a different *country*?

Is summer your favorite *season*?

Is a young cat called a *kitten*?

At University, are you a Psychology *student*?

3 Is a piano a musical *instrument*?

Is your favorite Jane Austen novel Pride and *Prejudice*?

| | | |
|---|---|---|
| | | Do you celebrate Christmas on the twenty fifth of *December*? |
| | | Is a trumpet a musical *instrument*? |
| | | Is Theresa May the prime *minister*? |
| | | Do you celebrate New Years eve on the thirty first of *December*? |
| | | Are dogs your favorite *animal*? |
| | | Do you like studying in the *library*? |
| Unpredictable | 1 | Do you often skip *lunch*? |
| | | Is your favorite food *fish*? |
| | | Do you have a sore *foot*? |
| | | Do most people have two *eyes*? |
| | | Have you been sightseeing in *France*? |
| | | Do you spent a lot of time on your *own*? |
| | | Do you spend a lot of money on *beer*? |
| | | Would you like to go running in the *rain*? |
| | | Do you have four *pets*? |
| | | Do you own a *house*? |
| | | Do you enjoy going to the *gym*? |
| | | Have you ever had to apologise to your *Dad*? |
| | | Do you have more than three *friends*? |
| | | Do you ever go to the *pub*? |
| | 2 | Have you ever watched a game of *cricket*? |
| | | Have you ever injured your *finger*? |

Is an apple the same colour as a *cherry*?

Have you ever played a game of *scrabble*?

Have you ever won a game of *poker*?

Do you have a good relationship with your *father*?

Are you allergic to *peanuts*?

Do you think you are good at *singing*?

Do you have two *siblings*?

Have you ever seen a wild *lion*?

Is an orange the same colour as a *carrot*?

Do you need a new *passport*?

Have you ever drawn a picture of a *person*?

3 At the weekend, did you do something nice for your *family*?

Do you know how to cook *spaghetti*?

Do you often see your *family*?

Have you ever visited the city of *Manchester*?

Do you want to buy a new *computer*?

Do you have a big *family*?

In your opinion, are you bad at *listening*?

Have you ever seen a big *elephant*?

## 7.3. Experimental materials used in Study 3

Table A5: List of stimuli used in Experiments 7-9.

| Question Predictability | Question | Answer Plausibility | Answer |
|---|---|---|---|
| Constraining | As well as cheese and tomato, which two toppings are usually on a Hawaiian pizza? | Plausible | Ham and pineapple |
| | | Implausible | December twenty fifth |
| Unconstraining | What would you like for dinner? | Plausible | Ham and pineapple |
| | | Implausible | December twenty fifth |
| Constraining | At which train station will you find platform nine and three quarters? | Plausible | Kings Cross |
| | | Implausible | It hit an iceberg |
| Unconstraining | Where are you getting a train from? | Plausible | Kings Cross |
| | | Implausible | It hit an iceberg |
| Constraining | How did The Titanic sink? | Plausible | It hit an iceberg |
| | | Implausible | Andy Murray |

| | | | |
|---|---|---|---|
| Unconstraining | What happened to your boat? | Plausible | It hit an iceberg |
| | | Implausible | Andy Murray |
| Constraining | What is Aurora Borealis commonly known as? | Plausible | The Northern Lights |
| | | Implausible | Ham and pineapple |
| Unconstraining | What can you see out of your window? | Plausible | The Northern Lights |
| | | Implausible | Ham and pineapple |
| Constraining | How often does the dentist tell you to brush your teeth? | Plausible | Twice a day |
| | | Implausible | Big Ben |
| Unconstraining | How often do you go outside for a walk? | Plausible | Twice a day |
| | | Implausible | Big Ben |
| Constraining | What is London's underground railway also known as? | Plausible | The Tube |
| | | Implausible | Hillary Clinton |
| Unconstraining | What is your least favorite method of transport? | Plausible | The Tube |

| | | | |
|---|---|---|---|
| | | Implausible | Hillary Clinton |
| Constraining | What are the names of Ron Weasley’s mum and dad? | Plausible | Molly and Arthur |
| | | Implausible | The Tube |
| Unconstraining | What are your parents called? | Plausible | Molly and Arthur |
| | | Implausible | The Tube |
| Constraining | What is the longest river in the world? | Plausible | The Amazon River |
| | | Implausible | Snow White |
| Unconstraining | Where did you go swimming yesterday? | Plausible | The Amazon River |
| | | Implausible | Snow White |
| Constraining | What colors are pandas? | Plausible | Black and white |
| | | Implausible | Tom Hanks |
| Unconstraining | What colors should I paint the wall? | Plausible | Black and white |
| | | Implausible | Tom Hanks |
| Constraining | What is the name of the British prime minister? | Plausible | Theresa May |
| | | Implausible | New York |

| | | | |
|---|---|---|---|
| Unconstraining | Who did you see when you visited London? | Plausible | Theresa May |
| | | Implausible | New York |
| Constraining | Which cutlery should I use to cut my food? | Plausible | A knife and fork |
| | | Implausible | Theresa May |
| Unconstraining | What did you buy from the shop? | Plausible | A knife and fork |
| | | Implausible | Theresa May |
| Constraining | What is the thirty first of December? | Plausible | New year’s eve |
| | | Implausible | Harry Potter |
| Unconstraining | When would you like to go for drinks? | Plausible | New year’s eve |
| | | Implausible | Harry Potter |
| Constraining | Which young wizard defeated Lord Voldemort? | Plausible | Harry Potter |
| | | Implausible | The White House |
| Unconstraining | What is the name of your favorite book? | Plausible | Harry Potter |
| | | Implausible | The White House |
| Constraining | Which famous clock is in London? | Plausible | Big Ben |

| | | | |
|---|---|---|---|
| | | Implausible | Twice a day |
| Unconstraining | What is your brother’s nickname? | Plausible | Big Ben |
| | | Implausible | Twice a day |
| Constraining | Who leads a gang of outlaws in Sherwood Forest? | Plausible | Robin Hood |
| | | Implausible | Ten Downing Street |
| Unconstraining | What is your best friend called? | Plausible | Robin Hood |
| | | Implausible | Ten Downing Street |
| Constraining | When do you celebrate Christmas? | Plausible | December twenty fifth |
| | | Implausible | A knife and fork |
| Unconstraining | When is your birthday? | Plausible | December twenty fifth |
| | | Implausible | A knife and fork |
| Constraining | Which character starred in the famous 007 films? | Plausible | James Bond |
| | | Implausible | Kings Cross |
| Unconstraining | What is your favorite film? | Plausible | James Bond |
| | | Implausible | Kings Cross |

| | | | |
|---|---|---|---|
| Constraining | When do you celebrate Halloween? | Plausible | October thirty first |
| | | Implausible | Robin Hood |
| Unconstraining | When do you next have a day off work? | Plausible | October thirty first |
| | | Implausible | Robin Hood |
| Constraining | Which tall building is in Paris? | Plausible | The Eiffel Tower |
| | | Implausible | October thirty first |
| Unconstraining | Where are you going at Christmas? | Plausible | The Eiffel Tower |
| | | Implausible | October thirty first |
| Constraining | Which river runs through London? | Plausible | The Thames |
| | | Implausible | Donald Trump |
| Unconstraining | Where did you go on your boat ride yesterday? | Plausible | The Thames |
| | | Implausible | Donald Trump |
| Constraining | Where does Father Christmas live? | Plausible | The North Pole |
| | | Implausible | New year’s eve |

| | | | |
|---|---|---|---|
| Unconstraining | Where would you like to go on holiday? | Plausible | The North Pole |
| | | Implausible | New year’s eve |
| Constraining | Where does the president of America live? | Plausible | The White House |
| | | Implausible | Molly and Arthur |
| Unconstraining | Where would you like to go when you visit America? | Plausible | The White House |
| | | Implausible | Molly and Arthur |
| Constraining | Which city is the statue of Liberty in? | Plausible | New York |
| | | Implausible | Buzz Lightyear |
| Unconstraining | Where would you like to go shopping? | Plausible | New York |
| | | Implausible | Buzz Lightyear |
| Constraining | Where does the prime minister live? | Plausible | Ten Downing Street |
| | | Implausible | James Bond |
| Unconstraining | Where would you like to go today? | Plausible | Ten Downing Street |

| | | | |
|---|---|---|---|
| | | Implausible | James Bond |
| Constraining | Which female candidate recently ran for president of the United States? | Plausible | Hillary Clinton |
| | | Implausible | The Northern Lights |
| Unconstraining | Who did you see when you visited America? | Plausible | Hillary Clinton |
| | | Implausible | The Northern Lights |
| Constraining | Where does the Queen live? | Plausible | Buckingham Palace |
| | | Implausible | Black and white |
| Unconstraining | Which tourist attraction would you like to visit in London? | Plausible | Buckingham Palace |
| | | Implausible | Black and white |
| Constraining | Which fictional character lived with seven dwarves? | Plausible | Snow White |
| | | Implausible | The Thames |
| Unconstraining | Who is your favorite fictional character? | Plausible | Snow White |
| | | Implausible | The Thames |

| | | | |
|---|---|---|---|
| Constraining | Who is the best Scottish tennis player? | Plausible | Andy Murray |
| | | Implausible | The North Pole |
| Unconstraining | Who is your favorite sportsman? | Plausible | Andy Murray |
| | | Implausible | The North Pole |
| Constraining | Who is the newly elected president of America? | Plausible | Donald Trump |
| | | Implausible | The Amazon River |
| Unconstraining | Who did you see interviewed on television recently? | Plausible | Donald Trump |
| | | Implausible | The Amazon River |
| Constraining | Which space ranger starred in Toy Story? | Plausible | Buzz Lightyear |
| | | Implausible | The Eiffel Tower |
| Unconstraining | Who is your favorite animated character? | Plausible | Buzz Lightyear |
| | | Implausible | The Eiffel Tower |

# 8. Appendix B: Linear mixed effects model outputs for the response time analysis of Experiments 1-6

Table B1: Linear mixed effects model output for the analysis of response times in Experiments 1-4. RE var = Random effects variance; (p) stands for random effects by participants; (i) stands for random effects by items. All predictors are defined in the Data Analysis section for each experiment.

| | | Experiment 1 | | |
|---|---|---|---|---|
| Predictor | Coeff | SE | t | RE var |
| Intercept | -136.21 | 55.53 | -2.45 | (p) 87806 (i) 13859 |
| Question Duration | -152.17 | 15.04 | -10.12 | - |
| Answer | - | - | - | - |
| Answer Agreement | - | - | - | - |
| Content | -28.31 | 29.00 | -0.97 | (p) 3498 |
| Length | -19.25 | 34.00 | -0.57 | (p) 10589 |
| Content * Length | -8.57 | 50.15 | -0.17 | (p) 0.00 |
| | | Experiment 2 | | |
| Predictor | Coeff | SE | t | RE var |
| Intercept | 380.26 | 57.60 | 6.60 | (p) 93329 (i) 18687 |

| | | | | |
|---|---|---|---|---|
| Question Duration | -72.88 | 17.25 | -4.23 | - |
| Answer | -21.86 | 16.46 | -1.33 | - |
| Answer Agreement | -55.21 | 15.17 | -3.64 | - |
| Content | -153.01 | 34.08 | -4.49 | (p) 5909 |
| Length | 10.89 | 33.25 | 0.33 | (p) 1171 |
| Content * Length | -110.21 | 63.75 | -1.73 | (p) 17340 |

Experiment 3

| Predictor | Coeff | SE | t | RE var |
|---|---|---|---|---|
| Intercept | -117.23 | 51.58 | -2.27 | (p) 71600 (i) 19603 |
| Question Duration | -124.99 | 13.92 | -8.98 | - |
| Answer | - | - | - | - |
| Answer Agreement | - | - | - | - |
| Content | 0.39 | 35.60 | 0.01 | (p) 0 |
| Length | 18.75 | 41.74 | 0.45 | (p) 3064 |
| Content * Length | - | - | - | - |

Experiment 4

| Predictor | Coeff | SE | t | RE var |
|---|---|---|---|---|
| Intercept | 483.65 | 60.19 | 8.04 | (p) 101914 (i) 12201 |

|  |  |  |  |  |
|---|---|---|---|---|
| Question Duration | -59.85 | 15.67 | -3.82 | - |
| Answer | -143.43 | 19.92 | -7.20 | - |
| Answer Agreement | -35.81 | 15.66 | -2.29 | - |
| Content | -81.68 | 39.07 | -2.09 | (p) 54 |
| Length | 28.19 | 36.81 | 0.77 | (p) 0 |
| Content * Length | - | - | - | - |

Table B2: Fixed (top) and random (bottom) effects structure for the linear mixed effects analysis of response times from final word onset (left) and final word offset (right) in Experiment 5. All predictors are defined in the Data Analysis section.

|  | Answer times from final word onset |  |  | Answer times from final word offset |  |  |
|---|---|---|---|---|---|---|
| Fixed effect | Coefficient | SE | t | Coefficient | SE | t |
| Intercept | 956.60 | 40.66 | 23.53 | 581.53 | 40.55 | 14.34 |
| Question Duration | -17.14 | 11.66 | -1.47 | -34.16 | 10.88 | -3.14 |
| Answer | -73.90 | 10.18 | -7.26 | -73.81 | 10.06 | -7.34 |
| Answer Agreement | -27.77 | 8.70 | -3.19 | -39.40 | 7.91 | -4.98 |
| Context Rate | -42.17 | 18.90 | -2.23 | -65.54 | 18.00 | -3.64 |
| Final Word Rate | -122.39 | 13.77 | -8.89 | 116.32 | 13.84 | 8.40 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Context Rate * Final Word Rate | 8.60 | 18.91 | 0.46 | 8.42 | 19.11 | 0.44 |
| Random effect | | Variance | *SD* | | Variance | *SD* |
| Item (Intercept) | | 7027 | 83.83 | | 5401 | 73.49 |
| Item (Context Rate) | | 2701 | 51.97 | | 2777 | 52.69 |
| Item (Final Word Rate) | | 1649 | 40.61 | | 1688 | 41.09 |
| Item (Context Rate * Final Word Rate) | | 6404 | 80.03 | | 6637 | 81.47 |
| Participant (Intercept) | | 50478 | 224.67 | | 50626 | 225.00 |
| Participant (Context Rate) | | 0 | 0.00 | | 0 | 0.00 |
| Participant (Final Word Rate) | | 2119 | 46.03 | | 2320 | 48.17 |
| Participant (Context Rate * Final Word Rate) | | 254 | 15.93 | | 438 | 20.93 |

Table B3: Fixed (top) and random (bottom) effects structure for the linear mixed effects analysis of response times from final word onset (left) and final word offset (right) in Experiment 6. All predictors are defined in the Data Analysis section.

| | Answer times from final word onset | | | Answer times from final word offset | | |
|---|---|---|---|---|---|---|
| Fixed effect | Coefficient | SE | t | Coefficient | SE | t |
| Intercept | 821.23 | 41.08 | 19.99 | 466.80 | 40.55 | 11.51 |
| Question Duration | -25.44 | 16.01 | -1.59 | -31.70 | 14.04 | -2.26 |
| Answer | -92.00 | 21.44 | -4.29 | -90.29 | 20.95 | -4.31 |
| Answer Agreement | -46.51 | 15.89 | -2.93 | -59.56 | 13.99 | -4.26 |
| Content Predictability | -201.81 | 41.18 | -4.90 | -153.46 | 38.03 | -4.04 |
| Final Word Rate | -125.79 | 19.56 | -6.43 | 106.70 | 19.41 | 5.50 |
| Syllable length | 15.59 | 14.51 | 1.07 | -13.99 | 12.71 | -1.01 |
| Content Predictability * Final Word Rate | 38.50 | 30.54 | 1.26 | 15.92 | 30.45 | 0.52 |
| Content Predictability * Syllable length | -35.79 | 28.84 | -1.24 | -20.36 | 25.50 | -0.80 |
| Final Word Rate * Syllable length | -23.83 | 15.71 | -1.52 | -5.72 | 15.71 | -0.36 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Content Predictability * Final Word Rate * Syllable length | 44.05 | 30.67 | 1.44 | 36.36 | 30.57 | 1.19 |
| Random effect | | Variance | *SD* | | Variance | *SD* |
| Item (Intercept) | | 99111 | 100 | | 6675 | 81.70 |
| Item (Final Word Rate) | | 0 | 0.00 | | 0 | 0.00 |
| Participant (Intercept) | | 47190 | 217.23 | | 47319 | 217.53 |
| Participant (Content Predictability) | | 19971 | 141.32 | | 19903 | 141.08 |
| Participant (Final Word Rate) | | 3053 | 55.25 | | 3302 | 57.46 |
| Participant (Syllable length) | | 0 | 0.00 | | 0 | 0.00 |
| Participant (Content Predictability * Final Word Rate) | | 0 | 0.00 | | 0 | 0.00 |
| Participant (Content | | 897 | 29.95 | | 1045 | 32.32 |

|  |  |  |  |  |
|---|---|---|---|---|
| Predictability * Syllable length) |  |  |  |  |
| Participant (Final Word Rate * Syllable length) | 335 | 18.31 | 381 | 19.52 |
| Participant (Content Predictability * Final Word Rate * Syllable length) | 0 | 0.00 | 0 | 0.00 |

Table B4: Random effects structure for the generalized linear mixed effects analysis of word report scores in Experiments 7, 8, and 9. All predictors are defined in the Data Analysis section.

|  | Experiment 7 |  | Experiment 8 |  | Experiment 9 |  |
|---|---|---|---|---|---|---|
| Random effect | Variance | *SD* | Variance | *SD* | Variance | *SD* |
| Item (Intercept) | 3.11 | 1.76 | 2.15 | 1.47 | 2.99 | 1.73 |
| Item (Question Predictability) | 0.33 | 0.57 | 0.07 | 0.27 | 0.88 | 0.94 |
| Item (Answer Plausibility) | 0.57 | 0.76 | 0.03 | 0.17 | 0.08 | 0.29 |
| Item (Block) | 0.11 | 0.33 | 0.17 | 0.41 | 0.18 | 0.43 |
| Item (Question Predictability * Answer Plausibility) | 0.20 | 0.44 | 0.03 | 0.19 | 0.02 | 0.15 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Item (Question Predictability * Block) | 0.36 | 0.60 | 0.03 | 0.17 | 0.45 | 0.68 |
| Item (Answer Plausibility * Block) | 0.41 | 0.64 | 0.02 | 0.15 | 0.02 | 0.15 |
| Item (Question Predictability * Answer Plausibility * Block) | 0.58 | 0.76 | 0.13 | 0.36 | 0.08 | 0.28 |
| Participant (Intercept) | 5.16 | 2.27 | 1.10 | 1.05 | 0.70 | 0.84 |
| Participant (Block) | 0.87 | 0.93 | 0.33 | 0.58 | 0.22 | 0.47 |

# 9. Appendix C: Bayesian mixed model outputs for the precision analysis of Experiments 1-6

Table C1: Model output for precision analyses in Experiments 1-4. Estimates are on the log scale (linear estimates in-text). (f) = fixed effect, (p) = RE by participants, (i) = RE by items.

| (Exp. 1) Predictor[a] | Estimate | SE | CrIs | Effective Sample |
|---|---|---|---|---|
| Intercept | (f) -0.82; | (f) 0.14; | (f) -1.08, -0.54; | (f) 347; |
|  | (p) 0.72; | (p) 0.11; | (p) 0.54, 0.97; | (p) 873; |
|  | (i) 0.41 | (i) 0.04 | (i) 0.34, 0.49 | (i) 1283 |
| shape_Intercept | (f) 0.40; | (f) 0.07; | (f) 0.27, 0.53; | (f) 629; |
|  | (p) 0.34; | (p) 0.05; | (p) 0.26, 0.45; | (p) 932; |
|  | (i) 0.23 | (i) 0.02 | (i) 0.34, 0.49 | (i) 1318 |
| Duration | (f) 0.17 | (f) 0.05 | (f) 0.07, 0.27 | (f) 1647 |
| shape_Duration | (f) -0.15 | (f) 0.03 | (f) -0.21, 0.09 | (f) 1698 |
| Content | (f) -0.03; | (f) 0.10; | (f) -0.22, 0.16; | (f) 1384; |
|  | (p) 0.18 | (p) 0.06 | (p) 0.05, 0.31 | (p) 544 |
| shape_Content | (f) 0.00; | (f) 0.06; | (f) -0.11, 0.10; | (f) 1612; |
|  | (p) 0.11 | (p) 0.04 | (p) 0.02, 0.19 | (p) 944 |
| Length | (f) -0.04; | (f) 0.11; | (f) -0.25, 0.17; | (f) 1617; |
|  | (p) 0.27 | (p) 0.06 | (p) 0.16, 0.40 | (p) 1577 |
| shape_Length | (f) -0.07; | (f) 0.06; | (f) -0.19, 0.04; | (f) 1645; |
|  | (p) 0.06 | (p) 0.04 | (p) 0.00, 0.13 | (p) 1004 |

| | | | | |
|---|---|---|---|---|
| Content * Length | (f) -0.24; | (f) 0.18; | (f) -0.60, 0.10; | (f) 1560; |
| | (p) 0.25 | (p) 0.12 | (p) 0.02, 0.50 | (p) 831 |
| shape_Content * Length | (f) -0.04; | (f) 0.10; | (f) -0.23, 0.16; | (f) 1785; |
| | (p) 0.07 | (p) 0.06 | (p) 0.00, 0.21 | (p) 1578 |

| (Exp. 2) Predictor | Estimate | SE | CrIs | Effective Sample |
|---|---|---|---|---|
| Intercept | (f) –0.44; | (f) 0.07; | (f) -0.58, 0.30; | (f) 443; |
| | (p) 0.33; | (p) 0.05; | (p) 0.15, 0.44; | (p) 569; |
| | (i) 0.25 | (i) 0.03 | (i) 0.20, 0.31 | (i) 1096 |
| shape_Intercept | (f) 0.19; | (f) 0.04; | (f) 0.10, 0.26; | (f) 571; |
| | (p); 0.19; | (p) 0.03; | (p) 0.14, 0.26; | (p) 954; |
| | (i) 0.06 | (i) 0.03 | (i) 0.00, 0.11 | (i) 469 |
| Duration | (f) 0.15 | (f) 0.04 | (f) 0.08, 0.22 | (f) 1405 |
| shape_Duration | (f) 0.04 | (f) -0.02 | (f) -0.07, 0.00 | (f) 2675 |
| Answer | (f) -0.01 | (f) 0.04 | (f) -0.10, 0.07 | (f) 3200 |
| shape_Answer | (f) 0.02 | (f) 0.03 | (f) -0.04, 0.07 | (f) 3200 |
| Agreement | (f) -0.06 | (f) 0.03 | (f) -0.13, 0.00 | (f) 1181 |
| shape_Agreement | (f) 0.02 | (f) 0.02 | (f) -0.01, 0.05 | (f) 2713 |
| Content | (f) 0.05; | (f) 0.12; | (f) -0.17, 0.28; | (f) 587; |
| | (p) 0.52 | (p) 0.08 | (p) 0.38, 0.71 | (p) 676 |
| shape_Content | (f) 0.10; | (f) 0.06; | (f) -0.02, 0.21; | (f) 797; |
| | (p) 0.25 | (p) 0.05 | (p) 0.16, 0.35 | (p) 1227 |
| Length | (f) 0.02; | (f) 0.08; | (f) -0.14, 0.18; | (f) 1281; |
| | (p) 0.21 | (p) 0.06 | (p) 0.10, 0.34 | (p) 1327 |
| shape_Length | (f) -0.01; | (f) 0.04; | (f) -0.08, 0.06; | (f) 2363; |
| | (p) 0.05 | (p) 0.04 | (p) 0.00, 0.13 | (p) 1231 |

| | | | | |
|---|---|---|---|---|
| Content*Length | (f) -0.19; | (f) 0.14; | (f) -0.48, 0.09, | (f) 1344; |
| | (p) 0.33 | (p) 0.13 | (p) 0.07, 0.58 | (p) 614 |
| shape_Content*Length | (f) 0.01; | (f) 0.08; | (f) -0.16, 0.17; | (f) 1576; |
| | (p) 0.26 | (p) 0.09 | (p) 0.07, 0.45 | (p) 720 |
| (Exp. 3) Predictor | Estimate | SE | CrIs | Effective Sample |
| Intercept | (f) -0.74; | (f) 0.14; | (f) -1.02, 0.47; | (f) 456; |
| | (p) 0.72, | (p) 0.10; | (p) 0.55, 0.94; | (p) 1032; |
| | (i) 0.49 | (i) 0.05 | (i) 0.41, 0.60 | (i) 1430 |
| shape_Intercept | (f) 0.44; | (f) 0.07; | (f) 0.31, 0.57; | (f) 876; |
| | (p) 0.31, | (p) 0.05; | (p) 0.24, 0.41; | (p) 1070; |
| | (i) 0.25 | (i) 0.03 | (i) 0.21, 0.31 | (i) 1697 |
| Duration | (f) 0.23 | (f) 0.05 | (f) 0.13, 0.32 | (f) 1739 |
| shape_Duration | (f) -0.10 | (f) 0.03 | (f) -0.16, -0.04 | (f) 2082 |
| Content | (f) 0.23; | (f) 0.15; | (f) -0.07, 0.53; | (f) 1292; |
| | (p) 0.28 | (p) 0.08 | (p) 0.11, 0.45 | (p) 802 |
| shape_Content | (f) 0.01; | (f) 0.08; | (f) -0.14, 0.17; | (f) 1327; |
| | (p) 0.10 | (p) 0.05 | (p) 0.01, 0.20 | (p) 802 |
| Length | (f) 0.10; | (f) 0.26; | (f) -0.41, 0.62; | (f) 1330; |
| | (p) 0.23 | (p) 0.14 | (p) 0.01, 0.54 | (p) 1010 |
| shape_Length | (f) 0.20; | (f) 0.14; | (f) -0.14, 0.17; | (f) 1596; |
| | (p) 0.18 | (p) 0.10 | (p) 0.01, 0.37 | (p) 993 |
| (Exp. 4) Predictor | Estimate | SE | CrIs | Effective Sample |
| Intercept | (f) -0.71; | (f) 0.11; | (f) -0.94, -0.50; | (f) 238; |
| | (p) 0.58; | (p) 0.08; | (p) 0.44, 0.76; | (p) 588; |
| | (i) 0.23 | (i) 0.03 | (i) 0.17, 0.30 | (i) 1279 |

| | | | | |
|---|---|---|---|---|
| shape_Intercept | (f) 0.28; | (f) 0.06; | (f) 0.16, 0.41; | (f) 335; |
| | (p) 0.58; | (p) 0.08; | (p) 0.24, 0.41; | (p) 677; |
| | (i) 0.10 | (i) 0.02 | (i) 0.06, 0.14 | (i) 1066 |
| Duration | (f) 0.11 | (f) 0.04 | (f) 0.04, 0.18 | (f) 1434 |
| shape_Duration | (f) 0.00 | (f) 0.02 | (f) -0.04, 0.04 | (f) 1955 |
| Answer | (f) 0.00 | (f) 0.05 | (f) -0.09, 0.09 | (f) 3200 |
| shape_Answer | (f) 0.13 | (f) 0.03 | (f) 0.06, 0.19 | (f) 3200 |
| Agreement | (f) 0.02 | (f) 0.04 | (f) -0.05, 0.09 | (f) 1689 |
| shape_Agreement | (f) 0.02 | (f) 0.02 | (f) -0.01, 0.06 | (f) 2090 |
| Content | (f) 0.12; | (f) 0.10; | (f) -0.08, 0.33; | (f) 1168; |
| | (p) 0.35 | (p) 0.08 | (p) 0.20, 0.52 | (p) 768 |
| shape_Content | (f) -0.17; | (f) 0.07; | (f) -0.30, -0.04; | (f) 1087; |
| | (p) 0.23 | (p) 0.05 | (p) 0.13, 0.35 | (p) 1194 |
| Length | (f) 0.01; | (f) 0.15; | (f) -0.28, 0.31; | (f) 1328; |
| | (p) 0.18 | (p) 0.12 | (p) 0.01, 0.44 | (p) 942 |
| shape_Length | (f) -0.08; | (f) 0.08; | (f) -0.25, 0.08; | (f) 2268; |
| | (p) 0.18 | (p) 0.12 | (p) 0.00, 0.25 | (p) 1508 |

[a] Models were fitted using a Weibull distribution (best fitting model assessed using LOO comparisons) and all predictors were the same as those in the lmer models. We ran 4 chains per model, each for 1600 iterations, with a burn-in period of 800, and initial parameter values set to zero. All of the reported models converged with no divergent transitions (all $\widehat{R}$ values $\leq 1.1$). Estimates are on the log scale. Note that if zero lies outside the credible interval (CrI), then we conclude there is sufficient evidence to suggest the estimate is different from zero. The shape parameter is most often used to model failure rates, and so is not relevant to the precision of responses.

The scale parameter quantifies the spread of the distribution (larger values of the scale parameter correspond to larger spread and less precise responses).

Table C2: Model output for precision analyses of answer times from final word onset (*italicized*) and final word offset (non-italicized) in Experiment 5. (f) = fixed effect, (p) = RE by participants, (i) = RE by items.

| Predictor[a] | Estimate | SE | CrIs | Effective |
|---|---|---|---|---|
| | | | | Sample |
| Intercept | *(f) -0.38;* | *(f) 0.07;* | *(f) -0.52, -0.24;* | *(f) 359;* |
| | *(p) 0.36;* | *(p) 0.05;* | *(p) 0.27, 0.48;* | *(p) 586;* |
| | *(i) 0.13* | *(i) 0.03* | *(i) 0.06, 0.18* | *(i) 847* |
| | (f) -0.35; | (f) 0.07; | (f) -0.48, -0.21; | (f) 282; |
| | (p) 0.38; | (p) 0.05; | (p) 0.29, 0.51; | (p) 771; |
| | (i) 0.12 | (i) 0.03 | (i) 0.06, 0.18 | (i) 708 |
| shape_Intercept | *(f) 0.33;* | *(f) 0.06;* | *(f) 0.23, 0.44;* | *(f) 372;* |
| | *(p) 0.29;* | *(p) 0.04;* | *(p) 0.22, 0.39;* | *(p) 878;* |
| | *(i) 0.14* | *(i) 0.02* | *(i) 0.10, 0.17* | *(i) 1415* |
| | (f) 0.36; | (f) 0.05; | (f) 0.26, 0.47; | (f) 418; |
| | (p) 0.28; | (p) 0.04; | (p) 0.22, 0.38; | (p) 681; |
| | (i) 0.09 | (i) 0.02 | (i) 0.04, 0.13 | (i) 564 |
| Question Duration | *(f) 0.16* | *(f) 0.03* | *(f) 0.10, 0.22* | *(f) 2327* |
| | (f) 0.18 | (f) 0.03 | (f) 0.12, 0.24 | (f) 2315 |
| shape_Question Duration | *(f) -0.03* | *(f) 0.03* | *(f) -0.08, 0.02* | *(f) 1776* |
| | (f) 0.00 | (f) 0.02 | (f) -0.04, 0.04 | (f) 2521 |
| Answer (No – Yes) | *(f) -0.04* | *(f) 0.04* | *(f) -0.11, 0.04* | *(f) 3200* |
| | (f) -0.03 | (f) 0.04 | (f) -0.10, 0.04 | (f) 3200 |
| shape_Answer | *(f) 0.11* | *(f) 0.03* | *(f) 0.06, 0.17* | *(f) 3200* |
| | (f) 0.07 | (f) 0.03 | (f) 0.02, 0.12 | (f) 3200 |
| Answer Agreement | *(f) -0.05* | *(f) 0.02* | *(f) -0.09, -0.01* | *(f) 3200* |

| | | | | |
|---|---|---|---|---|
| | (f) -0.02 | (f) 0.02 | (f) -0.06, 0.02 | (f) 3200 |
| shape_Answer Agreement | *(f) 0.03* | *(f) 0.02* | *(f) 0.00, 0.07* | *(f) 2423* |
| | (f) 0.03 | (f) 0.02 | (f) 0.00, 0.06 | (f) 3200 |
| Context Rate (Slow – Fast) | *(f) 0.27;* | *(f) 0.06;* | *(f) 0.15, 0.40;* | *(f) 1885;* |
| | *(p) 0.18;* | *(p) 0.06;* | *(p) 0.05, 0.30;* | *(p) 686;* |
| | *(i) 0.06* | *(i) 0.04* | *(i) 0.00, 0.17* | *(i) 1168* |
| | (f) 0.30; | (f) 0.07; | (f) 0.17, 0.43; | (f) 2069; |
| | (p) 0.19; | (p) 0.06; | (p) 0.06, 0.31; | (p) 1245; |
| | (i) 0.07 | (i) 0.05 | (i) 0.00, 0.18 | (i) 1076 |
| shape_Context Rate | *(f) 0.05;* | *(f) 0.05;* | *(f) -0.06, 0.16;* | *(f) 1669;* |
| | *(p) 0.14;* | *(p) 0.04;* | *(p) 0.06, 0.23;* | *(p) 932;* |
| | *(i) 0.16* | *(i) 0.05* | *(i) 0.05, 0.25* | *(i) 624* |
| | (f) 0.04; | (f) 0.05; | (f) -0.05, 0.13; | (f) 2072; |
| | (p) 0.12; | (p) 0.04; | (p) 0.02, 0.20; | (p) 680; |
| | (i) 0.09 | (i) 0.05 | (i) 0.00, 0.18 | (i) 577 |
| Final Word Rate (Slow – Fast) | *(f) 0.22;* | *(f) 0.08;* | *(f) 0.06, 0.37;* | *(f) 934;* |
| | *(p) 0.36;* | *(p) 0.06;* | *(p) 0.25, 0.50;* | *(p) 1108;* |
| | *(i) 0.16* | *(i) 0.07* | *(i) 0.02, 0.29* | *(i) 783* |
| | (f) -0.12; | (f) 0.10; | (f) -0.31, 0.07; | (f) 728; |
| | (p) 0.48; | (p) 0.08; | (p) 0.35, 0.65; | (p) 1245; |
| | (i) 0.10 | (i) 0.06 | (i) 0.00, 0.23 | (i) 763 |
| shape_Final Word Rate | *(f) 0.09;* | *(f) 0.04;* | *(f) 0.01, 0.18;* | *(f) 1933;* |
| | *(p) 0.18;* | *(p) 0.04;* | *(p) 0.10, 0.27;* | *(p) 1258;* |
| | *(i) 0.12* | *(i) 0.05* | *(i) 0.01, 0.21* | *(i) 556* |
| | (f) -0.13, | (f) 0.05; | (f) -0.24, -0.03; | (f) 1403; |
| | (p) 0.24; | (p) 0.04; | (p) 0.16, 0.33; | (p) 1127; |
| | (i) 0.09 | (i) 0.05 | (i) 0.01, 0.19 | (i) 492 |
| Context Rate * Final Word Rate | *(f) 0.04;* | *(f) 0.07;* | *(f) -0.11, 0.19;* | *(f) 3200;* |
| | *(p) 0.12;* | *(p) 0.09;* | *(p) 0.01, 0.32;* | *(p) 1315;* |
| | *(i) 0.12* | *(i) 0.09* | *(i) 0.01, 0.31* | *(i) 1235* |

|  |  |  |  |  |
|---|---|---|---|---|
|  | (f) 0.05;<br>(p) 0.21;<br>(i) 0.17 | (f) 0.08;<br>(p) 0.11;<br>(i) 0.11 | (f) -0.21, 0.11;<br>(p) 0.02, 0.42;<br>(i) 0.01, 0.41 | (f) 3200;<br>(p) 984;<br>(i) 994 |
| shape_Context Rate * Final Word Rate | *(f) -0.05;*<br>*(p) 0.14;*<br>*(i) 0.24* | *(f) 0.06;*<br>*(p) 0.08;*<br>*(i) 0.10* | *(f) -0.18, 0.06;*<br>*(p) 0.01, 0.30;*<br>*(i) 0.03, 0.43* | *(f) 3200;*<br>*(p) 608;*<br>*(i) 584* |
|  | (f) 0.09;<br>(p) 0.28;<br>(i) 0.19 | (f) 0.08;<br>(p) 0.08;<br>(i) 0.10 | (f) -0.06, 0.24;<br>(p) 0.12, 0.45;<br>(i) 0.01, 0.38 | (f) 2167;<br>(p) 851;<br>(i) 814 |

[a] Models were fitted using a Weibull distribution (best fitting model assessed using LOO comparisons) and all predictors were the same as those in the lmer models. We ran 4 chains per model, each for 1600 iterations, with a burn-in period of 800, and initial parameter values set to zero. All of the reported models converged with no divergent transitions (all $\widehat{R}$ values ≤ 1.1). Estimates are on the log scale. Note that if zero lies outside the credible interval (CrI), then we conclude there is sufficient evidence to suggest the estimate is different from zero. The shape parameter is most often used to model failure rates, and so is not relevant to the precision of responses. The scale parameter quantifies the spread of the distribution (larger values of the scale parameter correspond to larger spread and less precise responses).

Table C3: Model output for precision analyses of answer times from final word onset (*italicized*) and final word offset (non-italicized) in Experiment 6. (f) = fixed effect, (p) = RE by participants, (i) = RE by items.

| Predictor [a] | Estimate | SE | CrIs | Effective |
|---|---|---|---|---|
| | | | | Sample |
| Intercept | *(f) -0.64;* | *(f) 0.10;* | *(f) -0.84, -0.44;* | *(f) 555;* |
| | *(p) 0.50;* | *(p) 0.07;* | *(p) 0.38, 0.66;* | *(p) 1101;* |
| | *(i) 0.22* | *(i) 0.04* | *(i) 0.15, 0.29* | *(i) 1555* |
| | (f) -0.66; | (f) 0.10; | (f) -0.84, -0.47; | (f) 658; |
| | (p) 0.50; | (p) 0.07; | (p) 0.37, 0.66; | (p) 945; |
| | (i) 0.23 | (i) 0.04 | (i) 0.17, 0.31 | (i) 1494 |
| shape_Intercept | *(f) 0.31;* | *(f) 0.04;* | *(f) 0.23, 0.38;* | *(f) 1865;* |
| | *(p) 0.16;* | *(p) 0.03;* | *(p) 0.11, 0.23;* | *(p) 1457;* |
| | *(i) 0.13* | *(i) 0.03* | *(i) 0.08, 0.18* | *(i) 1465* |
| | (f) 0.28; | (f) 0.04; | (f) 0.19, 0.37; | (f) 1083; |
| | (p) 0.20; | (p) 0.03; | (p) 0.15, 0.28; | (p) 1292; |
| | (i) 0.10 | (i) 0.03 | (i) 0.05, 0.15 | (i) 1061 |
| Question Duration | *(f) 0.17* | *(f) 0.04* | *(f) 0.09, 0.26* | *(f) 2701* |
| | (f) 0.18 | (f) 0.04 | (f) 0.10, 0.26 | (f) 2161 |
| shape_Question Duration | *(f) -0.04* | *(f) 0.03* | *(f) -0.09, 0.01* | *(f) 3200* |
| | (f) -0.02 | (f) 0.02 | (f) -0.06, 0.03 | (f) 3200 |
| Answer (No – Yes) | *(f) -0.23* | *(f) 0.07* | *(f) -0.36, -0.10* | *(f) 3200* |
| | (f) -0.25 | (f) 0.06 | (f) -0.38, -0.12 | (f) 3200 |
| shape_Answer | *(f) 0.09* | *(f) 0.04* | *(f) 0.01, 0.18* | *(f) 3200* |
| | (f) 0.10 | (f) 0.04 | (f) 0.02, 0.19 | (f) 3200 |
| Answer Agreement | *(f) -0.02* | *(f) 0.04* | *(f) -0.10, 0.06* | *(f) 2649* |
| | (f) 0.00 | (f) 0.04 | (f) -0.08, 0.08 | (f) 2270 |
| shape_Answer Agreement | *(f) 0.01* | *(f) 0.03* | *(f) -0.04, 0.06* | *(f) 3200* |
| | (f) 0.00 | (f) 0.02 | (f) -0.05, 0.05 | (f) 3200 |
| Content Predictability | *(f) 0.18;* | *(f) 0.14;* | *(f) -0.10, 0.46;* | *(f) 1246;* |
| (Unpredictable - Predictable) | *(p) 0.64* | *(p) 0.10* | *(p) 0.47, 0.88* | *(p) 1169* |

| | | | | |
|---|---|---|---|---|
| | (f) 0.18;<br>(p) 0.61 | (f) 0.14;<br>(p) 0.10 | (f) -0.09, 0.46;<br>(p) 0.44, 0.83 | (f) 1114;<br>(p) 1249 |
| shape_Content Predictability | *(f) 0.19;*<br>*(p) 0.26* | *(f) 0.07;*<br>*(p) 0.06* | *(f) 0.05, 0.34;*<br>*(p) 0.17, 0.38* | *(f) 2119;*<br>*(p) 1607* |
| | (f) 0.11;<br>(p) 0.16 | (f) 0.06;<br>(p) 0.06 | (f) -0.01, 0.22;<br>(p) 0.03, 0.27 | (f) 2822;<br>(p) 696 |
| Final Word Rate (Slow – Fast) | *(f) -0.07;*<br>*(p) 0.21;*<br>*(i) 0.17* | *(f) 0.07;*<br>*(p) 0.08;*<br>*(i) 0.09* | *(f) -0.20, 0.07;*<br>*(p) 0.05, 0.36;*<br>*(i) 0.01, 0.35* | *(f) 3200;*<br>*(p) 791;*<br>*(i) 709* |
| | (f) -0.20;<br>(p) 0.28;<br>(i) 0.15 | (f) 0.08;<br>(p) 0.07;<br>(i) 0.15 | (f) -0.35, -0.06;<br>(p) 0.14, 0.43;<br>(i) 0.01, 0.32 | (f) 2431;<br>(p) 1253;<br>(i) 771 |
| shape_Final Word Rate | *(f) 0.08;*<br>*(p) 0.17;*<br>*(i) 0.14* | *(f) 0.05;*<br>*(p) 0.05;*<br>*(i) 0.06* | *(f) -0.02, 0.18;*<br>*(p) 0.05, 0.27;*<br>*(i) 0.01, 0.26* | *(f) 3182;*<br>*(p) 719;*<br>*(i) 482* |
| | (f) -0.11;<br>(p) 0.14;<br>(i) 0.17 | (f) 0.05;<br>(p) 0.06;<br>(i) 0.06 | (f) -0.20, -0.01;<br>(p) 0.03, 0.25;<br>(i) 0.04, 0.28 | (f) 3200;<br>(p) 794;<br>(i) 702 |
| Syllable length | *(f) 0.03;*<br>*(p) 0.04* | *(f) 0.04;*<br>*(p) 0.03* | *(f) -0.05, 0.10;*<br>*(p) 0.00, 0.10* | *(f) 2668;*<br>*(p) 1696* |
| | (f) 0.04;<br>(p) 0.04 | (f) 0.04;<br>(p) 0.03 | (f) -0.04, 0.11;<br>(p) 0.00, 0.11 | (f) 2323;<br>(p) 1186 |
| shape_Syllable length | *(f) -0.01;*<br>*(p) 0.04* | *(f) 0.02;*<br>*(p) 0.02* | *(f) -0.06, 0.04;*<br>*(p) 0.00, 0.09* | *(f) 3200;*<br>*(p) 1201* |
| | (f) -0.01;<br>(p) 0.02 | (f) 0.02;<br>(p) 0.02 | (f) -0.05, 0.03;<br>(p) 0.00, 0.06 | (f) 3200;<br>(p) 2151 |
| Content Predictability * Final Word Rate | *(f) 0.36;*<br>*(p) 0.16* | *(f) 0.11;*<br>*(p) 0.12* | *(f) 0.14, 0.58;*<br>*(p) 0.01, 0.44* | *(f) 3200;*<br>*(p) 1460* |
| | (f) -0.25;<br>(p) 0.15 | (f) 0.11;<br>(p) 0.11 | (f) -0.47, -0.04;<br>(p) 0.01, 0.40 | (f) 3200;<br>(p) 1480 |
| shape_Content Predictability * Final Word Rate | *(f) 0.12;*<br>*(p) 0.12* | *(f) 0.08;*<br>*(p) 0.09* | *(f) -0.03, 0.28;*<br>*(p) 0.00, 0.32* | *(f) 3200;*<br>*(p) 1411* |

| | | | | |
|---|---|---|---|---|
| | (f) -0.13;<br>(p) 0.09 | (f) 0.08;<br>(p) 0.07 | (f) -0.03, 0.04;<br>(p) 0.00, 0.26 | (f) 3200;<br>(p) 1642 |
| Content Predictability * Syllable length | *(f) 0.04;*<br>*(p) 0.13* | *(f) 0.08;*<br>*(p) 0.07* | *(f) -0.11, 0.20;*<br>*(p) 0.01, 0.29* | *(f) 2781;*<br>*(p) 873* |
| | (f) 0.15;<br>(p) 0.13 | (f) 0.08;<br>(p) 0.08 | (f) -0.01, 0.30;<br>(p) 0.00, 0.32 | (f) 2709;<br>(p) 803 |
| shape_Content Predictability * Syllable length | *(f) -0.03;*<br>*(p) 0.07* | *(f) 0.05;*<br>*(p) 0.05* | *(f) -0.13, 0.07;*<br>*(p) 0.00, 0.18* | *(f) 3200;*<br>*(p) 1187* |
| | (f) 0.02;<br>(p) 0.09 | (f) 0.05;<br>(p) 0.05 | (f) -0.07, 0.12;<br>(p) 0.01, 0.20 | (f) 3200;<br>(p) 902 |
| Final Word Rate * Syllable length | *(f) 0.03;*<br>*(p) 0.07* | *(f) 0.05;*<br>*(p) 0.05* | *(f) -0.07, 0.14;*<br>*(p) 0.00, 0.19* | *(f) 3200;*<br>*1827* |
| | (f) 0.00;<br>(p) 0.09 | (f) 0.06;<br>(p) 0.06 | (f) -0.02, 0.21;<br>(p) 0.00, 0.22 | (f) 3200;<br>(p) 1336 |
| shape_Final Word Rate * Syllable length | *(f) -0.08;*<br>*(p) 0.10* | *(f) 0.04;*<br>*(p) 0.05* | *(f) -0.16, 0.01;*<br>*(p) 0.01, 0.21* | *(f) 3200;*<br>*(p) 839* |
| | (f) 0.00;<br>(p) 0.09 | (f) 0.04;<br>(p) 0.05 | (f) -0.09, 0.08;<br>(p) 0.00, 0.20 | (f) 3200;<br>(p) 829 |
| Content Predictability * Final Word Rate * Syllable length | *(f) 0.21;*<br>*(p) 0.12* | *(f) 0.11;*<br>*(p) 0.09* | *(f) 0.00, 0.43;*<br>*(p) 0.00, 0.33* | *(f) 3200;*<br>*(p) 2128* |
| | (f) 0.11;<br>(p) 0.11 | (f) 0.10;<br>(p) 0.09 | (f) -0.09, 0.31;<br>(p) 0.00, 0.32 | (f) 3200;<br>(p) 2217 |
| shape_Content Predictability * Final Word Rate * Syllable length | *(f) -0.08;*<br>*(p) 0.12* | *(f) 0.08;*<br>*(p) 0.08* | *(f) -0.24, 0.07;*<br>*(p) 0.01, 0.31* | *(f) 3200;*<br>*(p) 1336* |
| | (f) -0.17;<br>(p) 0.11 | (f) 0.08;<br>(p) 0.08 | (f) -0.33, -0.01;<br>(p) 0.00, 0.30 | (f) 3200;<br>(p) 1269 |

[a] Models were fitted using a Weibull distribution (best fitting model assessed using LOO comparisons) and all predictors were the same as those in the lmer models. We ran 4 chains per model, each for 1600 iterations, with a burn-in period of 800, and initial parameter values set to zero. All of the reported models converged with no

divergent transitions (all $\widehat{R}$ values ≤ 1.1). Estimates are on the log scale. Note that if zero lies outside the credible interval (CrI), then we conclude there is sufficient evidence to suggest the estimate is different from zero. The shape parameter is most often used to model failure rates, and so is not relevant to the precision of responses. The scale parameter quantifies the spread of the distribution (larger values of the scale parameter correspond to larger spread and less precise responses