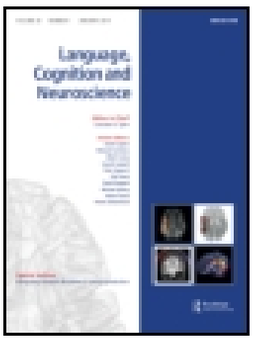

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Positional biases in predictive processing of intonation

Timo B. Roettger , Michael Franke & Jennifer Cole

To cite this article: Timo B. Roettger , Michael Franke & Jennifer Cole (2020): Positional biases in predictive processing of intonation, Language, Cognition and Neuroscience

To link to this article: https://doi.org/10.1080/23273798.2020.1853185

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2020.1853185

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Positional biases in predictive processing of intonation

Timo B. Roettger [a], Michael Franke[a] and Jennifer Cole[b]

[a]Institute of Cognitive Science, University of Osnabrück, Osnabruck, Germany; [b]Department of Linguistics, Northwestern University, Evanston, IL, USA

ABSTRACT

Real-time speech comprehension is challenging because communicatively relevant information is distributed throughout the entire utterance. In five mouse tracking experiments on German and American English, we probe if listeners, in principle, use non-local, early intonational information to anticipate upcoming referents. Listeners had to select a speaker-intended referent with their mouse guided by intonational cues, allowing them to anticipate their decision by moving their hand toward the referent prior to lexical disambiguation. While German listeners (Exps. 1–3) seemed to ignore early pitch cues, American English listeners (Exps. 4–5) were in principle able to use these early pitch cues to anticipate upcoming referents. However, many listeners showed no indication of doing so. These results suggest that there are important positional asymmetries in the way intonational information is integrated, with early information being paid less attention to than later cues in the utterance. Open data, scripts, and materials can be retrieved here: https://osf.io/xf8be/.

ARTICLE HISTORY
Received 28 July 2020
Accepted 9 November 2020

KEYWORDS
Prosody; intonation; sentence comprehension; rational analysis; mouse tracking

1. Introduction

Human speech is a complex communication signal that allows us to encode many different levels of meaning simultaneously. Beyond expressing propositional meaning, speakers use rhythmic and melodic aspects of speech to express pragmatic, social, and indexical meanings. Most notably, speakers modulate fundamental frequency (corresponding to what we perceive as pitch) to signal communicative functions (e.g. Cruttenden, 1997; Cutler et al., 1997; Dahan, 2015; Gussenhoven, 2004; Ladd, 2008; among many others). We henceforth refer to utterance-wide pitch modulation expressing non-propositional meaning as "intonation".

Speakers can pronounce the same sentence with different intonation contours. Different intonational events, described as local tonal configurations, can occur in many different positions, thereby indicating different interpretations (Pierrehumbert & Hirschberg, 1990). This represents a challenge to listeners as they need to know what parts of the intonation contour to attend to and how to update the discourse model in relation to the information gleaned from the intonation contour. Nevertheless, in most situations, listeners experience no problems interpreting intonation. Despite the obvious importance of intonation for many languages, we know surprisingly little about how listeners integrate information from intonation for the purpose of identifying the meaning the speaker intended to convey. In five mouse tracking experiments on German and American English, we shed light on these questions and investigate what parts of an intonation contour listeners use to predict the utterance meaning.

1.1. Listeners integrate intonational information rationally

Recognising the pragmatic, social or indexical meaning that a speaker intends to convey through the content and form of their utterance is an error-prone process because speech transmission is often imperfect. An intended message is often only partially received (e.g. Jaeger, 2010; Levy & Jaeger, 2007) and the information that is received is often of limited reliability (e.g. Pogue et al., 2016). How do language users cope with this high level of uncertainty?

One possible answer to this question is rooted in how language users process information. Increasingly, models of human cognition (e.g. Anderson, 1990; Geisler, 2011; Knill & Richards, 1996), and language processing in particular, propose that humans are rational in how they integrate information and adapt to new environments (e.g. Franke & Jäger, 2016; Goodman & Frank, 2016; Kleinschmidt & Jaeger, 2015). These

CONTACT Timo B. Roettger timo.b.roettger@gmail.com Institute of Cognitive Science, University of Osnabrück, Osnabruck, 49069, Germany
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2020.1853185.

models draw on the idea that people take uncertainty and noisy signals into account and perform a given task in a statistically optimal way. Rational models of language processing have been successfully applied to speech perception and adaptation (e.g. Feldman et al., 2009; Kleinschmidt & Jaeger, 2015), pragmatic reasoning (e.g. Degen & Tanenhaus, 2015; Franke, 2009; Frank & Goodman, 2012; Franke & Jäger, 2016; Goodman & Frank, 2016; Schuster & Degen, 2020), and, crucially, intonation and intonational processing (e.g. Bergen & Goodman, 2015; Buxó-Lugo & Kurumada, 2019; Roettger & Franke, 2019).

Here we assume a model of a rational comprehender who rapidly integrates information in the speech signal in order to probabilistically predict likely upcoming linguistic information. In that sense, information integration of a rational comprehender has the following properties that are rooted in independent evidence: First, information integration is incremental (e.g. Grodner et al., 2010; Kamide et al., 2003), i.e. information is integrated as soon as it becomes available. Second, information integration is predictive (e.g. Crocker, 2010; Levy, 2008), i.e. possible continuations of an observed partial utterance are anticipated. And third, information integration is probabilistic (e.g. Kuperberg & Jaeger, 2016; Spivey-Knowlton et al., 1993), i.e. multiple potential interpretations are activated at the same time and may receive different weights of plausibility at any moment in time.

For example, several studies have demonstrated that the mapping of referential expressions to a corresponding visual stimulus is initiated on the basis of incomplete, partial information before the disambiguating lexical item is fully available, including information about the referential context (e.g. Degen & Tanenhaus, 2015; Grodner et al., 2010; Spivey et al., 2002; Tanenhaus et al., 1995), presuppositions and affordances relevant to intended actions (e.g. Chambers et al., 2004), and differences between their own perspective and that of their interlocutors (Hanna et al., 2003; Heller et al., 2008; Nadig & Sedivy, 2002). Intonation is no exception. Comprehenders rapidly integrate intonational cues to anticipate a likely speaker-intended referent even before disambiguating lexical material is heard (e.g. Dahan et al., 2002; Ito & Speer, 2008; Kurumada et al., 2014a; Roettger & Rimland, 2020; Weber et al., 2006). For example, West Germanic languages express discourse relevant functions by the position of pitch accents, i.e. intonational events co-occurring with lexically stressed syllables (e.g. Ladd, 2008). In German and English, for instance, the position and form of a pitch accent can signal a referent as discourse-given, or contrastive (e.g. Grice & Baumann, 2007; Pierrehumbert & Hirschberg, 1990). For example, a constituent with a high rising pitch movement can signal contrastive information as in example (1), contrasting the sentence *object* with a set of alternatives (e.g. pear) and in (2), contrasting the sentence *subject* with a set of alternatives (e.g. Heinrich).

(1) Margarethe ate an APPLE.
→ Margarethe did not eat a pear.
(2) MARGARETHE ate an apple.
→ It was not Heinrich who ate an apple.

Listeners use this knowledge to predict the discourse status of the current referent, and of upcoming referents, as soon as they hear reliable intonational cues. For instance, Kurumada et al. (2014a) showed that a high rising accent on the verb "look" in "It LOOKS like a zebra" induces anticipatory eye-movements to a picture that contains a referent that looks like the zebra but is slightly different. Thus, listeners rapidly integrated a conventionalised intonational cue to anticipate an upcoming referent (here, an accent on "looks" in the construction "it looks like X" implies the referent is not actually X). It is, of course, rational to use conventionalised cues since they are informative for comprehending intended meaning. Additionally, listeners can change their expectations about the reliability of these cues. They can both learn to predict upcoming referents based on less-conventionalised cues and unlearn conventionalised cues.

For example, Roettger and Franke (2019) showed that German listeners can use both the presence of a pitch accent on the verb and its absence to predict the discourse status of an upcoming referent (see also Morett & Fraundorf, 2019 for similar findings on beat gestures and pitch accents). In their experiments with German, listeners first heard a discourse-setting question introducing a referent (3a) and then heard an answer to this question, either confirming the already mentioned referent (3b) or contrasting the mentioned referent with a new one (3c). Within the microcosm of the experiment, the discourse status of the target referent as given (Geige*/violin*) or contrastive (Birne*/pear*) could be systematically identified by the presence or absence of an early pitch accent on the auxiliary verb (hat*/has*). The authors showed that listeners use these cues to anticipate the referent. Some of these patterns only emerged after listeners had encountered a sufficient amount of evidence to learn the association between intonation contour and meaning. Moreover, listeners' anticipatory use of the absence of an early pitch accent was substantially slower than what was observed for the presence of an early pitch accent. The authors

argued that an asymmetry in cue reliability led to these asymmetric anticipation patterns.

(3) a. Hat der Wuggy dann die Geige aufgesammelt?
Did the wuggy pick up the violin?
b. Der Wuggy hat dann die Geige aufgesammelt.
The wuggy then picked up the violin.
c. Der Wuggy hat dann die Birne aufgesammelt.
The wuggy then picked up the pear.

In order to formalise their findings, Roettger and Franke (2019) introduce a computational model in which a rational comprehender considers the predictive value of an intonational cue for different possible utterance interpretations. The predictive value of a cue depends on two sources of information: The prior reliability of a mapping between that cue and a communicative function (based on listeners' experience with their language) and the likelihood of this mapping given the listeners' most recent experiences. This model captures listeners' adaptation behaviour: If listeners are exposed to stimuli in which an otherwise informative cue is unreliable, they downgrade the informational value of that cue (see also Kurumada et al., 2014b; Roettger & Rimland, 2020). Conversely, if listeners are exposed to stimuli in which an otherwise unreliable cue is very informative, they selectively upgrade its informational value. This model relates a part of the speech signal (here components of an intonation contour) to its reliability for comprehending speaker-intended intonational meaning. Roettger and Franke's model does not, in principle, need to refer to particular structural primitives (e.g. a pitch accent) that listeners identify to predict meaning (e.g. Pierrehumbert & Hirschberg, 1990). Instead, what matters is the amount of disambiguating information carried by any given realised (partial) contour. In other words, the proposed model posits rational comprehenders who use whatever information is available to them in a probabilistically optimal way, whether that is a traditionally assumed intonational feature (like a pitch accent) or not.

The proposed predictive processing of acoustic information is in line with research on speech rate normalisation and distal prosodic effects. There is evidence suggesting that listeners' expectation of upcoming information is dependent on temporal properties of previously encountered speech, affecting phoneme monitoring (e.g. Cutler, 1976; Rysling et al., 2020), category boundaries between temporally defined speech sound categories (e.g. Kidd, 1989; Reinisch & Sjerps, 2013), lexical stress (Reinisch et al., 2011), and the perception of function words (Baese-Berk et al., 2014; Dilley & Pitt, 2010). These studies suggests that listeners track relevant prosodic information in order to process upcoming information.

Thus, a rational approach to prosodic processing seems to be a useful point of departure for understanding how listeners successfully integrate intonation to infer speaker intentions in light of the ubiquitous variability associated with these aspects of human speech (e.g. Bolinger, 1972; Cole, 2015; Grice et al., 2017; Hirschberg, 2002).

1.2. Positional biases – prenuclear vs. nuclear tunes

An important limitation of the aforementioned work is that it has primarily looked at one particular component of an intonation contour, usually the right-most pitch accent (or absence thereof) in a sentence or intonational phrase. However, an intonation contour may span a large temporal window and can have multiple prominent pitch events distributed throughout the utterance. Successful models of intonational processing should be able to predict how different aspects of this complex signal are integrated. This is particularly important as not all pitch events are equal in function or perceived prominence. Traditionally, a special functional status is assigned to the last pitch accent in a phrase and the following boundary tone (the "nuclear contour", e.g. Halliday, 1967, Cruttenden, 1997). This focus on the nuclear contour is partly due to the belief that the intonation signal preceding the nuclear contour, i.e. the prenuclear contour, is not relevant for expressing discourse meaning. In (3b-c) above, for example, "the Wuggy" might also carry an optional prenuclear pitch accent preceding the nuclear pitch accent on "violin" or "pear", which does not necessarily modulate the interpretation of the utterance. Pitch accents temporally preceding the rightmost pitch accent in a phrase are henceforth referred to as prenuclear accents.

The relevance of prenuclear accents for the expression of discourse meaning is empirically unclear. Prenuclear accents in English have been described as optional and variable in production (e.g. Chodroff & Cole, 2018) and not reliably marking information-structural distinctions (Gussenhoven, 2011, 2015). Prenuclear accents have been described as "ornamental" (Büring, 2007), and as placed for rhythmic purposes only (Calhoun, 2010). They also have been described as having lesser acoustic prominence than nuclear accents and are less likely than nuclear accents to be perceived as prominent by listeners (Cole et al., 2019). In artificial language learning experiments with English-speaking participants, these prenuclear components of the intonation contour are not paid attention

to as much as nuclear components (Kapatsinski et al., 2017). Taken together, these findings suggest that prenuclear accents are, at a minimum, less relevant for expressing speaker-intended meaning, justifying their minor role in the literature on meaning and intonation.

In contrast, there are several sources of evidence suggesting that early intonational cues in an utterance, including prenuclear pitch accents, can be communicatively relevant. Several studies have provided evidence that the tonal contour early on in the utterance can be probabilistically associated with different interpretations of referential and speech act meaning (e.g. Braun & Asano, 2013; Petrone & D'Imperio, 2011; Petrone & Niebuhr, 2014). For example, Petrone and Niebuhr (2014) showed that German listeners use the shape and alignment of prenuclear pitch accents to distinguish statements from questions in a gating experiment. In line with these perception findings, production studies indicate that the form of prenuclear accents in German is modulated in accordance with information structural distinctions (e.g. Braun & Asano, 2013; Féry & Kügler, 2008). Several authors found that the presence of prenuclear accents affected listeners' judgments about the congruence (Breen et al., 2010; Gussenhoven, 1983) and appropriateness (Birch & Clifton, 1995) of narrow versus broad focus contexts. For instance, the sentence in (1) (repeated as 4) is an appropriate answer to both questions (i) "What did Margarethe eat?" and (ii) "What did Margarethe do?".

(4) Margarethe ate an apple.

In the answer to the first question (i), only the verb complement *apple* is in (narrow) focus. In the answer to the second question (ii), the whole verb phrase is in (broad) focus. The verb can carry an optional prenuclear pitch accent which seems to be more appropriate for the broad focus context. However, results from the studies cited above also suggest that prenuclear accents play a lesser role for listeners' behaviour than nuclear accents. Following up on these studies, Bishop (2017) presented evidence from cross-modal priming studies suggesting that the presence of a prenuclear accent was compatible with a broad focus domain, but it disrupted priming for narrow focus contexts. Yet other work suggests that the semantic integration of prenuclear accents seems to be dependent on the pitch accent type and thus its perceptual saliency. For instance, Braun and Biezma (2019) showed that a rising-falling prenuclear pitch accent can activate semantic alternatives just like nuclear pitch accents do (e.g. Husband & Ferreira, 2016). However, they only found an effect for a very salient pitch accent type that is commonly used for contrastive interpretations, indicating that the type of a tonal event plays a role in whether and how listeners interpret parts of the early contour.

Given these two competing bodies of research, it becomes clear that we have not fully understood how listeners integrate different parts of intonation contours to comprehend speakers' intended meaning. While it is generally agreed that the nuclear part of the contour is informative for important discourse-pragmatic distinctions, for the prenuclear part of the contour the evidence is mixed. One body of research suggests that listeners ignore prenuclear accents and infer speaker intentions based on the nuclear portion of the intonation contour only. We refer to this position as the NUCLEAR-ONLY account. As opposed to that, there is substantial evidence suggesting that listeners can use prenuclear accents at least to some extent (which might be dependent on the type of prenuclear event). We refer to this position as the PRENUCLEAR-MATTERS account.

From the perspective of a rational comprehender, these two accounts make predictions about the listeners' prior beliefs about the usefulness of a prenuclear accent for successful intention recognition, i.e. listeners' expectations based on their experience with their language. The NUCLEAR-ONLY account predicts that prenuclear accents will be ignored, i.e. listeners' prior expectations about the usefulness of prenuclear accents is zero. The PRENUCLEAR-MATTERS account predicts that prenuclear accents can be informative for successful intention recognition to some extent, i.e. listeners' prior expectations about the usefulness of prenuclear accent is greater than zero and they should use these aspects of the signal to predict speaker intentions.

Regardless of listeners' prior beliefs in the usefulness of a cue, (HYPER)RATIONAL comprehenders should be able to adjust their prior beliefs about the usefulness of a cue in light of sufficient reliable evidence. Within an experiment that systematically presents a prenuclear accent with a particular discourse interpretation, we would expect hyper-rational listeners, modelled as hyper-rational agents in the tradition of classical economic theory (de Finetti, 1931; Savage, 1954; Von Neumann & Morgenstern, 1944), to adapt their expectations and learn to use this cue over the course of the experiment, a state of affairs that is supported for nuclear parts of the intonation contour (Roettger & Franke, 2019). On the other hand, language comprehenders might not be hyper-rational, but rather *adaptively rational* (e.g. Anderson, 1990; Chater & Oaksford, 1999, 2000; Gigerenzer & Goldstein, 1996; Hagen et al., 2012; Tversky & Kahneman, 1981). An adaptively rational

agent has evolved choice mechanisms, sensory and representational capacities that have evolved to work well in the kinds of situations that the agent has confronted most frequently or, more generally, that matter most to the agent's evolutionary fitness (e.g. Bednar & Page, 2007; Fawcett et al., 2013; Galeazzi & Franke, 2017; Hammerstein & Stevens, 2012; McNamara, 2013). If information in prenuclear accents is almost never informative for non-local meaning distinctions downstream, it might therefore make sense that an adaptively rational interpreter does not, as a general rule or heuristic, pay attention to this part of the input stream, at least not as much as to other parts of the incoming linguistic signal.

1.3. The present study

The present study attempts to answer two related questions. Do listeners use prenuclear accents to predict upcoming information (PRENUCLEAR-MATTERS) or not (NUCLEAR-ONLY), and does exposure to reliable mappings between prenuclear accents and pragmatic functions increase the extent to which listeners use these parts of the speech signal? We assess listeners' predictive behaviour and potential adaptation effects using the mouse tracking paradigm in a two-alternative forced choice task (e.g. Spivey et al., 2005). A large body of experiments has demonstrated that the continuous uptake of sensory input can be reflected in participants' hand or finger movements (e.g. Dotan et al., 2019; Freeman, 2018, for recent overviews). This has also been shown for intonational processing. Listeners integrate intonational information early on and move their mouse towards a likely target referent before they have processed disambiguating lexical information (Tomlinson et al., 2017). In our experiments, listeners hear question-answer pairs in which the answer refers to either an already mentioned (given) referent or a non-mentioned (contrastive) referent. Listeners are instructed to click on one of two visually presented referents.

We adopt the linking hypothesis that listeners' certainty about the interpretation of an utterance influences the final moment in time, relative to the unfolding speech signal, at which listeners' mouse movements turn towards the target referent that is eventually chosen. In other words, we assume that the more reliably a partial utterance indicates an interpretation, the earlier the listener will integrate it to anticipate the speaker's referential intentions. This assumption is in line with the general idea of ballistic accumulator models (e.g. Ratcliff & McKoon, 2008), according to which evidence in favour of a choice or hypothesis accumulates stochastically over time and results in execution if a critical mass is met.

In our first experiment, we replicate Roettger and Franke's results on the presence vs. absence of an early nuclear pitch accent in German in order to establish a temporal baseline (experiment 1). We then extend the paradigm to test how listeners' anticipatory behaviour compares to the perception of informative prenuclear pitch accents (experiment 2–3). We find no compelling evidence that prenuclear accents in German are used to anticipate speaker intentions. Nor do we find evidence that listeners can learn to use an early prenuclear pitch accent predictively. German, however, restricts us to some extent because it is less flexible with regard to the type of prenuclear pitch accent that it licenses in certain positions. In two follow-up experiments (experiments 4–5), we present American English listeners to highly informative and salient prenuclear pitch accents. It turns out that listeners can learn to use these early cues. However, we observe a large amount of variability across and within listeners.

2. Experiment 1: replicating Roettger and Franke (2019)

The following experimental set up is based on Roettger and Franke (2019) and follows their design very closely. We ask whether German listeners can anticipate speakers' referential intentions based on the presence vs. absence of a nuclear pitch accent on the verb.

2.1. Method

2.1.1. Participants and procedure

Thirty German speakers participated in the study. All participants had self-reported normal or corrected-to-normal vision and normal hearing (12 males, 18 females, mean age = 25 (SD = 3.6)).

Participants were seated in front of a Mac mini 2.5 GHz Intel Core i5. They controlled the experiment via a Logitech B100 corded USB Mouse. Cursor acceleration was linearised and cursor speed was slowed down (to 1400 sensitivity) using the CursorSense© application (version 1.32). Slowing down the cursor ensured that motor behaviour was recorded as the acoustic signal unfolded resulting in a smooth trajectory from start to target (Kieslich et al., 2019b).

Participants were told about a fantasy creature called "Wuggy" that carries things around. There were twelve different objects that the wuggy could pick up (bee, chicken, diaper, fork, marble, pants, pear, rose, saw, scale, vase, violin).

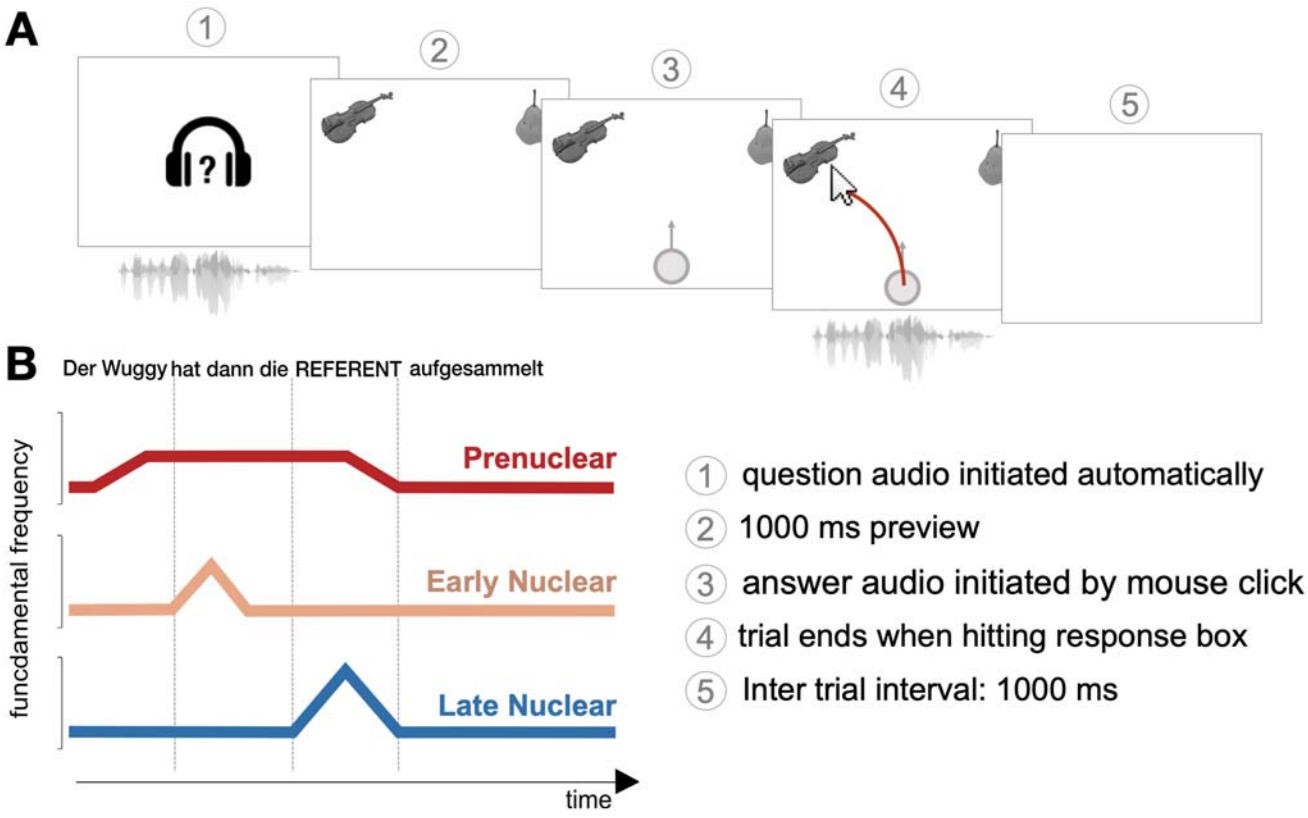

Figure 1. A – Schematic depiction of experimental trials in experiments 1–3. On the context screen (1), participants heard the context-setting question. After a 1000 ms preview of target and competitor referents (2), the initiation button is displayed at the bottom centre of the screen (3). Upon clicking the initiation button, listeners started the audio playback of the response sentence, indicating the target referent (4). The trial ended when hitting the response box around the target referent (5). Inter trial intervals were 1000 ms. B – Schematic depiction of intonation contours used in experiment 1–3. Note that the lexical condition (see text for description) exhibits the same intonation contour as the late nuclear condition.

Each trial exposed participants first to a context screen, which was shown for 2500 ms and provided a specific discourse context (see Figure 1A). The context screen displayed an uninformative image of a headphone. Participants heard either a topic question like (5), which introduced a referent as given in the discourse, or a neutral question like (6) introducing no specific discourse content:

(5) Hat der Wuggy dann die Geige aufgesammelt?
Did the wuggy then pick up the violin?
(6) Was ist passiert?
What happened?

Following the context screen, participants saw a response screen with two visually presented referents, each depicting one object in the upper left and right corner, respectively (left/right placement of target vs. competitor response alternatives was counterbalanced within participants and items). After 1000 ms, a yellow circle appeared at the bottom centre of the screen. When participants clicked on the yellow circle, they initiated playback of an audio recording of a statement specifying which object was picked up, e.g. Geige (Engl. *violin*) or Birne (Engl. *pear*):

(7) Der Wuggy hat dann die Geige aufgesammelt.
The wuggy has then the violin picked-up.
The wuggy then picked up the violin.
(8) Der Wuggy hat dann die Birne aufgesammelt.
The wuggy has then the pear picked-up.
The wuggy then picked up the pear.

There were three intonation contours (see Figure 1B) mapped onto three discourse contexts. After a neutral question (6), the entire sentence has broad focus (i.e. all constituents are discourse-new), which in German can be prosodically encoded by a high(rising) pitch accent in the default position, i.e. on the stressed syllable of the verb argument (blue contour in Figure 1B). In this discourse context, listeners need to wait for the lexical item of the object to know which picture to click on; we henceforth refer to this condition as the LEXICAL condition. An alternative intonation pattern is also

available for a German broad focus sentence, with a prenuclear rise in pitch on the subject, followed by high plateau and a nuclear fall in pitch preceding the referent. This pattern has been described as the hat pattern in past research (e.g. Ambrazaitis & Niebuhr, 2008; Braun, 2006, see red line in Figure 1B). Since we are interested in the position of the first potentially relevant intonational cue, we will henceforth refer to this contour as the PRENUCLEAR condition. This contour will be relevant for Exp. 2 and 3 but not for Exp. 1.

After a polar topic question (5), the utterance in (7) answers in the affirmative and identifies the object to be selected as the given referent. This confirmation can be prosodically encoded through a high rising accent on the auxiliary verb "hat" (Engl. *has*), as an instance of the verum focus construction (Engl: The Wuggy HAS picked up the violin.) We henceforth refer to this contour as the EARLY NUCLEAR condition (orange line in Figure 1B). Finally and in contrast to the verum focus construction, the answer in (8) negates the topic question (5). It affirms a different referent for the object than the one stated in the question, which would typically be expressed through a contrastive focus construction, characterised by an intonation contour with a high rising nuclear accent on the verb object Birne (Engl. *pear*). We henceforth refer to this contour as the LATE NUCLEAR condition (blue line in Figure 1B). Note again that the intonation contour in the LATE NUCLEAR condition is identical to the contour in the LEXICAL condition. All possible statements referring to a specific referent (n = 12) occurred either as an answer to a neutral question in the LEXICAL condition or as an affirmative or corrective answer to the topic question in the EARLY / LATE NUCLEAR condition, resulting in 36 different target sentences overall. Each target sentence was repeated four times, resulting in a total of 144 target trials (12 items * 3 conditions * 4 repetitions).

Participants were instructed to move the mouse immediately upwards after clicking the initiation button (see Spivey et al., 2005, Kieslich et al., 2019a) and to choose the picture that matches the object referent as quickly as possible (by moving into one of the response boxes, see Kieslich et al., 2019a). If they did not initiate their movement immediately within 350 ms, they automatically received feedback that reminded them to do so. This time pressure ensured that participants began their mouse movement (straight upward) before the onset of relevant acoustic information, which enables distinguishing properties in the acoustic signal to influence the continuous motor output during its movement. After each response selection, the screen was left blank for a 1000 ms inter trial interval. Visual inspection of all trajectories (by TR) ensured that participants generally followed these instructions.

Prior to the experimental trials, participants familiarised themselves with the paradigm during 16 practice trials. The combination of condition and target referent were pseudorandomized for each block, and the order of trials within a block and order of blocks were randomised for each participant.

2.1.2. Materials

Visual stimuli were taken from the BOSS corpus (Brodeur et al., 2010). There were two sets of acoustic stimuli: Questions (topic, neutral) providing a discourse context and answers (in statement form), produced in three intonation patterns, designed to trigger participants' mouse selection of the picture on the response screen that corresponds to the object of the verb.

Acoustic stimuli were recorded by two trained phoneticians in a sound-attenuated booth at the Institute of Phonetics in Cologne with a headset microphone (AKG C420) using 48 kHz/16-bit sampling. One male speaker produced the discourse-setting questions which were also used in previous studies (Roettger & Franke, 2019; Roettger & Rimland, 2020). Another male speaker produced natural answers congruent with the prompting question. The speaker produced three intonation contours: (i) A contour with an "early" nuclear accent on the verb (EARLY NUCLEAR); (ii) A contour with a "late" rising accent on the default position, which is the accented syllable of the sentence object (LATE NUCLEAR), here the critical referent. As described above, this contour is congruent both with an answer to the neutral question and with a corrective answer to the topic question; (iii) A hat pattern with a prenuclear rising accent on the subject and a falling accent on the object (PRENUCLEAR). The PRENUCLEAR contour is only used in Exps. 2 and 3 (see below).

To ensure that sentences across the three different intonation contours exhibit the same temporal characteristics (i.e. the lexical information of all words becomes available at the same time across conditions), sentences are manipulated and resynthesized using Praat (Boersma & Weenink, 2016) applying the procedure described in appendix A1 (the online repository contains both original and resynthesized stimuli at https://osf.io/xf8be/). Note that this procedure differs from the one employed in Roettger and Franke (2019), who resynthesized all contours based on one baseline utterance.

2.1.3. Data analysis

The x, y screen coordinates of the computer mouse were sampled at 100 Hz using the mousetrap plugin (Kieslich

& Henninger, 2017) implemented in the open source experimental software OpenSesame (Mathôt et al., 2012). Trajectories were processed with the package mousetrap (Kieslich & Henninger, 2017) using R (R Core Team, 2020).[1]

For each trial, we computed the following measurement based on space-normalized trajectories. We look at the moment in time relative to the unfolding speech signal at which a mouse trajectory starts to migrate uninterruptedly towards the target interpretation. We define the turn-towards-the-target (TTT) as the latest point in time at which the trajectory did not head towards the target (see Roettger & Franke, 2019).[2]

We fitted Bayesian hierarchical linear models to turn-towards-the-target measurements as a function of DISCOURSE context and centralised BLOCK, as well as their two-way interaction, using brms (Bürkner, 2016). The models included maximal random-effect structures, allowing the predictors and their interactions to vary by participants and experimental items (DISCOURSE x BLOCK). We used weakly informative Gaussian priors centred around zero with $\sigma = 250$ ms for all population-level regression coefficients (e.g. Gelman et al., 2008), truncated Gaussians priors for all standard deviations (mean = 0, sd = 100), and LKJ(2) priors for all correlation parameters.[3] Four sampling chains with 2000 iterations each were run for each model, with a warm-up period of 1000 iterations.

We report, for relevant predictor levels and differences between predictor levels, 95% credible intervals (CrIs). A 95% credible interval demarcates the range of values that comprise 95% of probability mass of our posterior beliefs. We also report the probability of direction (also known as the Maximum Probability of Effect – MPE, see Makowski et al., 2019). The probability of direction varies between 0.5 and 1 and can be interpreted as the probability that a parameter (described by its posterior distribution) is strictly positive or negative, i.e. is larger or smaller than 0. It is mathematically defined as the proportion of the posterior distribution that is of the median's sign. We consider estimates for which MPE is close to 1 as compelling evidence.

2.1.4. Predictions

In line with the rational comprehender model (Roettger & Franke, 2019), we expect that listeners can use both the presence and the absence of an early pitch accent on the verb indicating that the upcoming discourse referent is given or contrastive, respectively. In other words, listeners who hear the verb (either with or without a pitch accent), can already predict the discourse status of the upcoming referent. Thus, we expect compelling evidence that the difference between LEXICAL and both EARLY NUCLEAR and LATE NUCLEAR is different from a point null-hypothesis (i.e. no difference between groups, corresponding to the position of a sceptic). Moreover, a rational comprehender should integrate the EARLY NUCLEAR accent on the verb exhibiting high evidential strength substantially earlier than the cue corresponding to the absence of that accent exhibiting low evidential strength (the absence of a pitch accent on the verb in the LATE NUCLEAR condition). In other words, we predict responses to the EARLY NUCLEAR condition to be anticipated earlier than responses to the LATE NUCLEAR condition. This prediction follows directly from Roettger and Franke's rationale (2019) that, in this particular German construction, the absence of a pitch accent on the verb should be less reliably associated with the discourse status of the upcoming referent than the presence of a pitch accent on the verb. We further expect the weak LATE NUCLEAR cue to become integrated increasingly faster over the course of the experiment as listeners learn to use this cue given enough exposure.

2.2. Results

The whole data set of a participant was excluded whenever they (a) exhibited more than 10% errors, or (b) exhibited movement behaviour violating instructions in more than 15% of the trials, or (c) exhibited initiation times above 350 ms in more than 15% of the trials. We excluded one subject due to too much unwanted movement behaviour.

Trials with initiation times greater than 350 ms (0.4%) and incorrect responses (0.5%) were discarded on a trial-by-trial basis. Additionally, trials that exhibited movement behaviour violating instructions were discarded, too (1.9%). The remaining data were statistically analysed.

Because of the form of the trajectories (initially gravitating towards the middle and then smoothly turning towards the target) and due to the stimuli spanning large temporal windows, it is informative to investigate properties of the trajectories relative to temporal landmarks in the acoustic stimuli. We are interested in the question of when listeners' manual movements indicate that the available evidence in the acoustic signal makes the target referent more likely than the competitor. In the following, we will analyse said turn-towards-the-target measurement. Figure 2A displays the horizontal cursor position over time as a function of discourse context indicating clear temporal differences between conditions: These differences are already apparent at the point at which the cursor starts turning towards the target (i.e. lines go up in Figure 2A).

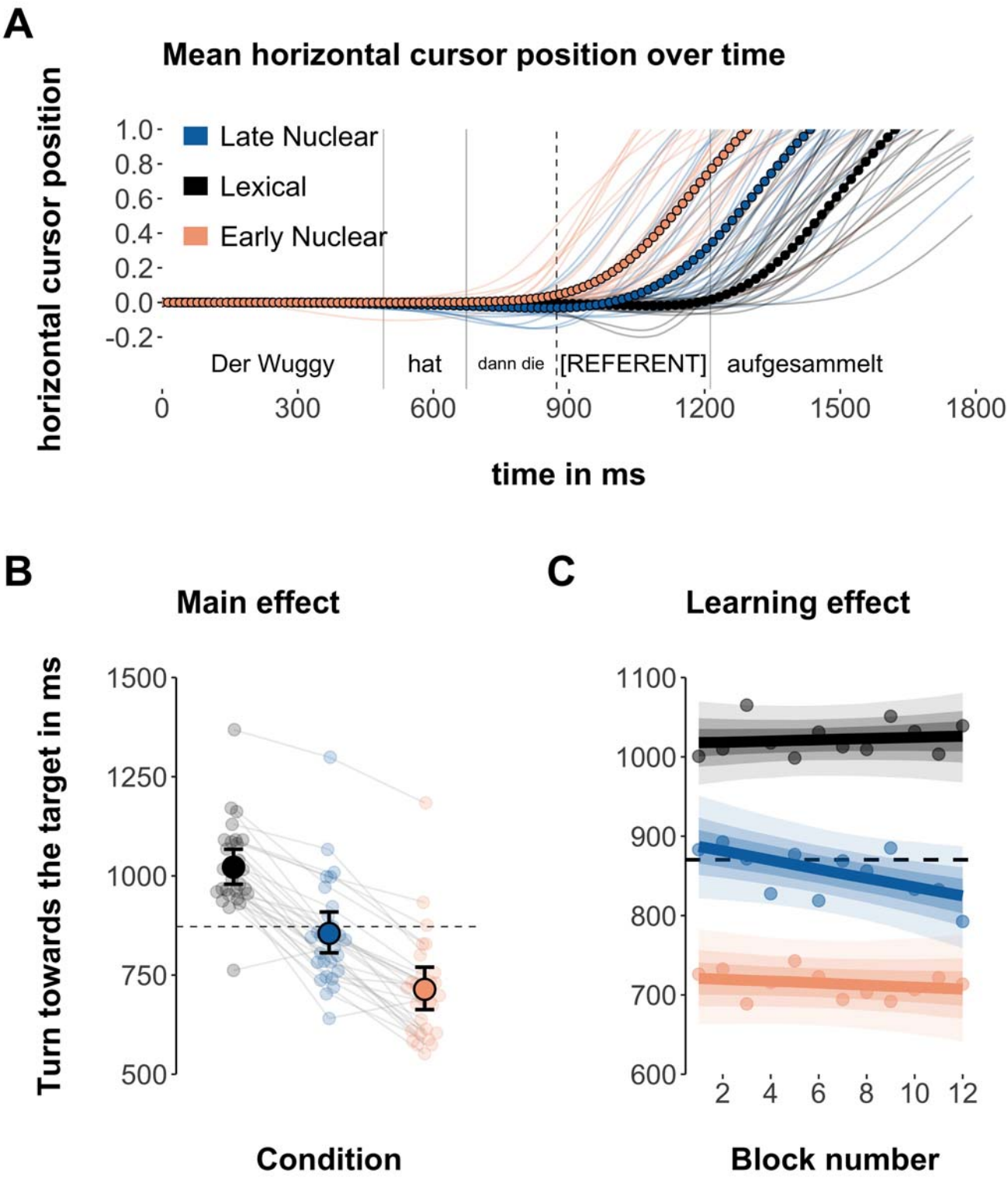

Figure 2. Results of experiment 1. A – Horizontal cursor position of space-normalized averaged trajectories (y) plotted against time of acoustic stimulus (x) (semi-transparent lines are average values for each participant); B – Posterior means and 95% credible intervals of turn-towards-the-target measures effects across conditions (semi-transparent points are average values for each participant); C – Posterior means and 50% / 75% / 95% credible intervals of turn-towards-the-target effects across conditions and experimental blocks (points indicate grand average values); black dashed line indicates the average acoustic onset of the referent.

Table 1. Posterior means, 95% credible intervals, as well as the posterior probability of direction of experiment 1.

Parameter (experiment 1)	posterior mean	95% CI	MPE
Difference: LEXICAL – EARLY NUCLEAR	307	(272;340)	~1.00
Difference: LEXICAL – LATE NUCLEAR	166	(132;201)	~1.00
Difference: EARLY – LATE NUCLEAR	−141	(−185; −106)	~1.00
Block effect: LEXICAL	2	(−17;21)	0.60
Block effect: EARLY NUCLEAR	−4	(−25;17)	0.65
Block effect: LATE NUCLEAR	−20	(−44;3)	0.95

Figure 2 and Table 1 display the main effects and learning effects of experiment 1. The LEXICAL condition serves as a baseline as listeners have to wait until the acoustic information about the referential expression becomes available. It takes listeners on average 1022 ms to start turning towards the target (β = 1022 [979,1067]). There is compelling evidence that the intonationally informed trials (EARLY and LATE NUCLEAR) elicit earlier TTT values than LEXICAL trials, with TTTs being earlier in the EARLY NUCLEAR condition (β = 714 [663,770]) than in LATE NUCLEAR trials (β = 855 [806,909]). This ordinal relationship (LEXICAL > LATE NUCLEAR > EARLY NUCLEAR) is rather consistent as can be seen in Figure 2B with the majority of participants showing the same ordinal relationship.

These temporal effects may change dynamically across the course of the experiment (see Figure 2C, Table 1). While neither LEXICAL nor EARLY NUCLEAR trials show a clear development over the course of the experiment (LEXICAL: β = 2 [−17,22], EARLY NUCLEAR: β = −4 [−25,17]), there is weak evidence that TTT measures become earlier throughout the experiment in the LATE NUCLEAR trials (β = −20 [−44,3]), i.e. the overwhelming majority of posterior samples is smaller than zero but zero remains a plausible value.

2.3. Discussion

The data of experiment 1 suggest that intonational information can facilitate referential intention recognition in the presence of relevant discourse information. This is in line with previous work (e.g. Dahan et al., 2002; Ito & Speer, 2008; Kurumada et al., 2014a; Weber et al., 2006;). More specifically, we replicated recent findings presented by Roettger and Franke (2019) who demonstrated these effects for the same German constructions. The acoustically early nuclear pitch accent on the verb allows listeners to anticipate an intended referent long before the lexical material becomes available.

Listeners also make use of prosodic information prior to the lexical referent in the LATE NUCLEAR trials, which are characterised by a flat intonation contour and a late high rising pitch accent on the sentence object. This pitch accent becomes acoustically available simultaneously with the lexical referent (thus temporally later than in the EARLY NUCLEAR condition). The fact that TTTs are earlier in the LATE NUCLEAR condition compared to the LEXICAL condition means that participants are making decisions about the referent before encountering the lexical or intonational cues on the sentence object. The early cue in the LATE NUCLEAR condition is the absence of an accent on the verb

In Roettger and Franke (2019), stimuli were resynthesized in such a way that the acoustic form of EARLY NUCLEAR trials and LATE NUCLEAR trials were identical except for the f0 contour. Based on their findings, the authors concluded that in LATE NUCLEAR trials, listeners must have used the absence of the pitch accent on the verb to anticipate the upcoming referent. The present study qualitatively replicates this finding with more natural stimuli that potentially contain prosodic cues distributed throughout the whole utterance. The strikingly similar patterns to Roettger and Franke (2019) suggest that listeners probably focus on the relevant pitch movements (or the absence thereof) despite having access to other cues prior to these landmark cues.

The anticipation in the LATE NUCLEAR condition does not happen as early as in the EARLY NUCLEAR condition but still has a temporal advantage over simple lexical disambiguation (LEXICAL > LATE NUCLEAR > EARLY NUCLEAR). Roettger and Franke (2019) have argued that this temporal difference can be derived from differences in listeners' initial beliefs about the predictive value of prosodic cues, with the absence of a pitch accent being a weak cue to a contrastive referent. This account assumes that the temporal differences between LATE NUCLEAR and EARLY NUCLEAR result from different evidential strengths associated with these different intonational cues which are based on prior experiences with the listeners' native language.

Taking these findings as a point of departure, Exps. 2 and 3 test listeners' integration of prenuclear pitch accents, i.e. pitch accents that temporally precede the rightmost pitch accent in a phrase. As opposed to Exp. 1, we included the above described PRENUCLEAR contour, characterised by a pitch rise on the sentence subject, followed by a high plateau and a fall in pitch towards the sentence object. Within the experiments, the PRENUCLEAR condition is either paired with the LATE NUCLEAR (Exp. 2) or the EARLY NUCLEAR condition (Exp. 3) in how it is associated with the alternative referential interpretation, respectively (see Table 2, i.e. a PRENUCLEAR contour indicates a contrastive referent when paired with EARLY NUCLEAR; it indicates a given referent

when paired with LATE NUCLEAR). The hat contour in the PRENUCLEAR condition commonly has a neutral meaning in German, usually presenting self-evident information. The contour has no strong association with the discourse status of one particular referent (but might affect the interpretation of referent relationships in, for example, "A or B"-constructions, see Ambrazaitis & Niebuhr, 2008). Both referential interpretations, one in which the sentence object is contrastive and one in which it is given, are valid interpretations for native speakers.

In the PRENUCLEAR condition, listeners have access to substantially earlier pitch accent information than in both the EARLY NUCLEAR and the LATE NUCLEAR condition (see Figure 1B above). In the PRENUCLEAR-MATTERS account, we expect the PRENUCLEAR trials to elicit earlier turn-towards-the-targets than in both LATE NUCLEAR and EARLY NUCLEAR trials. In the NUCLEAR-ONLY account, we expect listeners to ignore the prenuclear accent. Regardless of listeners' initial beliefs about form-function mappings, a (HYPER)RATIONAL comprehender account predicts that listeners will learn to predictively use the prenuclear accent if reliably co-occurring with a referential discourse relationship. Listeners should rapidly adjust their prior beliefs and learn to associate intonational cues with respective meaning. If true we would expect turn-towards-the-target values for PRENUCLEAR trials to decrease over the course of the experiment.

3. Experiment 2 and 3: prenuclear accents in German

3.1. Method

The method of experiment 2 and 3 differed only in the experimental stimuli from experiment 1. Below we specify the differences between experiment 1 and 2–3. Experiments 2 and 3 were preregistered on the 6th of March 2018 and 24th of April 2018 prior to data collection, respectively.

3.1.1. Participants and procedure

Thirty German speakers participated in each experiment. All participants had self-reported normal or corrected-to-normal vision and normal hearing (Exp. 2: 11 males, 19 females, mean age = 25.3 (SD = 3.5), Exp. 3: 15 males, 15 females, mean age = 25 (SD = 3.6)).

3.1.2. Materials

In addition to the PRENUCLEAR contour, we used the same materials as specified for Exp. 1, but intonation contours were paired up differently. In Exp. 2, listeners were exposed to PRENUCLEAR and EARLY NUCLEAR contours (alongside LEXICAL trials), with PRENUCLEAR contours being associated with contrastive referents, and EARLY NUCLEAR contours being associated with given referents. In Exp. 3, listeners were exposed to PRENUCLEAR and LATE NUCLEAR contours (alongside LEXICAL trials), with PRENUCLEAR contours being associated with given referents, and LATE NUCLEAR contours being associated with contrastive referents. This way, we were able to compare the PRENUCLEAR trials to both LATE NUCLEAR and EARLY NUCLEAR trials. Simultaneously, we could see whether PRENUCLEAR contours are biased towards given or contrastive interpretations.

3.1.3. Data analysis

All data were analysed as specified for experiment 1 in §2.1.3.

3.2. Results

Following preregistered exclusion protocol (same as in Exp. 1), trials with initiation times greater than 350 ms (Exp. 2: 0.5%, Exp. 3: 0.5%) and incorrect responses (Exp. 2: 0.2%, Exp. 3: 0.6%) were discarded on a trial-by-trial basis. Additionally, trials that exhibited movement behaviour violating instructions were discarded, too (Exp. 2: 1.2%, Exp. 3: 1.6%). We excluded one participant due to an error rate above 10% (Exp. 2) and one participant due to illicit movement behaviour in more than 15% of trials (Exp. 3). The remaining data were statistically analysed as specified above.

3.2.1. Results of experiment 2

Table 3 displays the results of experiment 2. There is compelling evidence that the intonationally informed conditions (EARLY NUCLEAR and PRENUCLEAR) elicit earlier TTT values than in LEXICAL trials, with TTTs being earlier in the EARLY NUCLEAR condition ($\beta = 755$ [717,799]) than in the PRENUCLEAR condition ($\beta = 873$ [825,922]).

We do not find any evidence that turn-towards-the-target measures change throughout the course of the experiment (see Figure 3, Table 3) (LEXICAL: $\beta = 5$ [−10,24], PRENUCLEAR: $\beta = -5$ [−24,15], EARLY NUCLEAR: $\beta = -9$ [−32,14]).

3.2.2. Results of experiment 3

Table 4 displays the results of Exp. 3. Despite highly overlapping posteriors with the intonationally informed conditions (LATE NUCLEAR: $\beta = 945$ [894,994], PRENUCLEAR: $\beta = 933$ [874,991]), there is substantial evidence that the differences between LEXICAL and both PRENUCLEAR and LATE NUCLEAR are greater than 0 (LEXICAL –

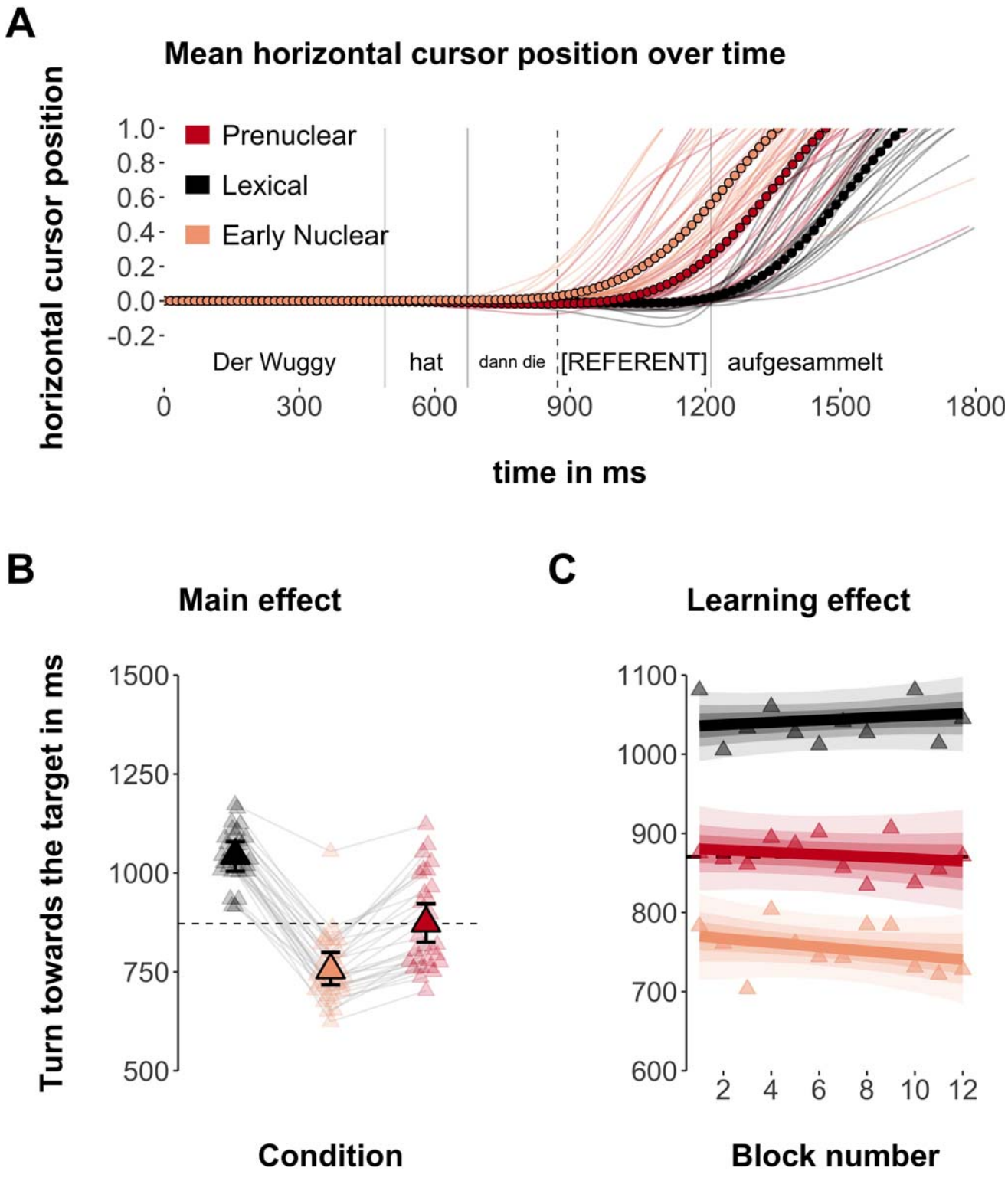

Figure 3. Results of experiment 2. A – Horizontal cursor position of space-normalized averaged trajectories (y) plotted against time of acoustic stimulus (x) (semi-transparent lines are average values for each participant); B – Posterior means and 95% credible intervals of turn-towards-the-target measures effects across conditions (semi-transparent points are average values for each participant); C – Posterior means and 50% / 75% / 95% credible intervals of turn-towards-the-target effects across conditions and experimental blocks (points indicate grand average values); the black dashed line indicates the average acoustic onset of the referent

Table 2. Crossing of intonation contour and discourse status across experiments 1–3.

	Hat der Wuggy dann die Geige aufgesammelt? *Did the wuggy then pick up the violin?* example sentence	intonation	discourse status
Exp 1	Der Wuggy **HAT** dann die Geige aufgesammelt.	**EARLY NUCLEAR**	given
	Der Wuggy hat dann die **BIRNE** aufgesammelt.	**LATE NUCLEAR**	contrastive
Exp 2	Der Wuggy **HAT** dann die Geige aufgesammelt.	**EARLY NUCLEAR**	given
	Der **WUGGY** hat dann die BIRNE aufgesammelt.	**PRENUCLEAR**	contrastive
Exp 3	Der **WUGGY** hat dann die Geige aufgesammelt.	**PRENUCLEAR**	given
	Der Wuggy hat dann die **BIRNE** aufgesammelt.	**LATE NUCLEAR**	contrastive

Table 3. Posterior means, 95% credible intervals, as well as the posterior probability of direction of experiment 2.

Parameter (experiment 2)	posterior mean	95% CI	MPE
Difference: LEXICAL – PRENUCLEAR	171	(129;218)	~1.00
Difference: LEXICAL – EARLY NUCLEAR	288	(256;320)	~1.00
Difference: PRENUCLEAR – EARLY NUCLEAR	117	(78;162)	~1.00
Block effect: LEXICAL	5	(−10;24)	0.72
Block effect: PRENUCLEAR	−5	(−24;15)	0.69
Block effect: EARLY NUCLEAR	−9	(−32;14)	0.79

PRENUCLEAR: $\beta = 76$ [38,111], LEXICAL – LATE NUCLEAR: $\beta = 63$ [33,95]). However, there is no compelling evidence for a reliable difference between PRENUCLEAR and LATE NUCLEAR ($\beta = -13$ [−51,29]). Again, there is no indication that the turn-towards-the-target values become smaller or larger throughout the experiment (see Figure 4, Table 4).

3.3. Omnibus analysis and discussion

Experiments 2 and 3 tested listeners' integration of intonational information in the PRENUCLEAR condition, characterised by a prenuclear pitch accent on the subject: Within the microcosm of the experiment, this contour offered a reliable cue to discourse disambiguation at an early point in the utterance, namely the sentence subject which acoustically unfolds a couple of hundred milliseconds before the verb. Interestingly, these intonation contours did not elicit earlier turn-towards-the-target values. Although arguably a salient perceptual cue, the rising pitch accent on the sentence subject was not immediately integrated to anticipate the intended referent, i.e. PRENUCLEAR trials systematically elicited later TTTs than the EARLY NUCLEAR condition. These findings suggest that, for the discourse meaning tested in this experiment, the NUCLEAR-ONLY account makes the correct predictions.

Table 4. Posterior means, 95% credible intervals, as well as the posterior probability of direction of experiment 3.

Parameter (experiment 3)	posterior mean	95% CI	MPE
Difference: LEXICAL – PRENUCLEAR	76	(38;111)	~1.00
Difference: LEXICAL – LATE NUCLEAR	63	(33;95)	~1.00
Difference: PRENUCLEAR – LATE NUCLEAR	−13	(−51;29)	0.74
Block effect: LEXICAL	7	(−12;25)	0.77
Block effect: PRENUCLEAR	−6	(−27;20)	0.71
Block effect: LATE NUCLEAR	2	(−19;23)	0.59

While the NUCLEAR-ONLY account did not expect listeners to use this early cue right from the beginning of the experiment (since this intonational pattern might not be strongly associated with the information structure of a subsequent referent), a (HYPER)RATIONAL comprehender should have learned this form-function association throughout exposure. However, there is no evidence for any dynamic adjustment over the course of the experiment.

If we qualitatively compare the PRENUCLEAR trials of Exp. 2 to the LATE NUCLEAR trials in Exp. 1, they elicit strikingly similar temporal patterns. In PRENUCLEAR trials of Exp. 2, listeners could potentially use the salient prenuclear pitch accent on the subject to anticipate the contrastive referent. Instead of integrating this early pitch event, listeners seem not to exploit this cue and end up performing similar to the LATE NUCLEAR condition in Exp. 1. This similarity suggests that listeners in Exp. 2 might merely pay attention to the absence of the nuclear accent on the verb.

In line with this interpretation, the anticipatory advantage of both LATE NUCLEAR trials and PRENUCLEAR trials in Exp. 3 almost vanishes because listeners do not have the comparison to the EARLY NUCLEAR condition, including the pitch accent on the verb. In Exp. 3, both LATE NUCLEAR and PRENUCLEAR conditions are only slightly faster than lexical disambiguation. To quantify said temporal disadvantage in Exp. 3 (and the uncertainty associated with this quantification), we fitted an omnibus model to the turn-towards-target measurements as a function of DISCOURSE context, scaled BLOCK number, and EXPERIMENT, as well as their three-way interaction. The model included maximal random-effect structures, allowing the predictors and their interactions to vary by participants (DISCOURSE x

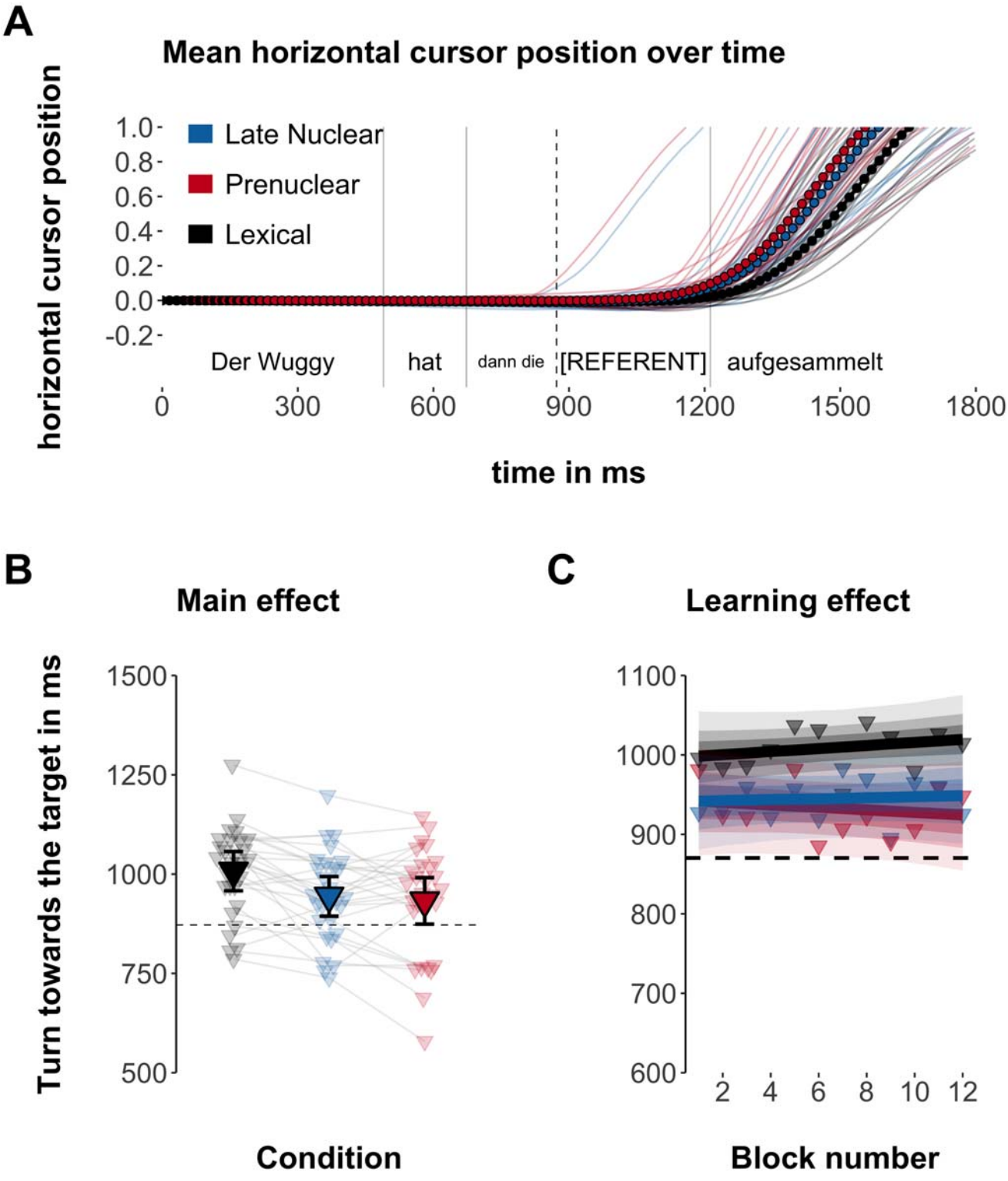

Figure 4. Results of experiment 3. A – Horizontal cursor position of space-normalized averaged trajectories (y) plotted against time of acoustic stimulus (x) (semi-transparent lines are average values for each participant); B – Posterior means and 95% credible intervals of turn-towards-the-target measures effects across conditions (semi-transparent points are average values for each participant); C – Posterior means and 50% / 75% / 95% credible intervals of turn-towards-the-target effects across conditions and experimental blocks (points indicate grand average values); the black dashed line indicates the average acoustic onset of the referent

BLOCK) and experimental items (DISCOURSE x BLOCK x EXPERIMENT).

The model suggests that there is compelling evidence that the LATE NUCLEAR condition of Exp. 3 (in the absence of the competing EARLY NUCLEAR condition) elicited slower turn-towards-the-target values than its counterpart in Exp. 1 (in the presence of the competing EARLY NUCLEAR contour) (Difference: LATE NUCLEAR (Exp. 1)–(Exp. 3): $\beta = -89$ [−151,–25]). There is also weak evidence that the PRENUCLEAR condition of Exp. 3 (in the absence of the competing EARLY NUCLEAR condition) elicited slower TTTs than its counterpart in Exp. 2 (Difference: PRENUCLEAR (Exp. 2)–(Exp. 3): $\beta = -58$ [−122,8]), i.e. the overwhelming majority of posterior samples is smaller than zero but zero is a plausible value (see Table 5).

In sum, our results suggest that within the limited microcosm of the experiment, listeners can predictively use the presence of a nuclear pitch accent on the verb and its absence (replicating Roettger & Franke, 2019). In the latter case, however, the absence of a pitch accent only yields an anticipatory advantage when listeners are exposed to an available alternative *with* the pitch accent (as in Exps. 1 and 2). This is an interesting finding as it suggests that temporal cue integration is not only contingent on the evidential strength of the cue (prior) and recently experienced mappings between cue and speaker intention (likelihood), but it is also contingent on the set of alternative cues. The presence of a highly informative cue on the verb in Exp. 1 and Exp. 2 seems to direct listeners attention to this particular time window. Paying attention to the verb enables them to predictively exploit even the absence of a pitch accent in this position. When such a highly informative cue on the verb is not present in the microcosm (e.g. in Exp. 3), listeners' information integration seems substantially slower. This pattern can be interpreted in two ways. Listeners do not attend to the time window associated with the verb and thus the absence of a pitch accent is treated as less informative and integrated in a delayed manner. Alternatively, listeners may indeed use the prenuclear pitch accent, but do so in an even more delayed way.

Coming back to our hypotheses, our data suggests that listeners either do not use prenuclear pitch accents to anticipate the referential status of the upcoming referent at all or they do so with a large time lag. Moreover, a (HYPER)RATIONAL comprehender approach does not explain our data well. Listeners received systematic evidence that the prenuclear accent co-occurs with a particular discourse interpretation, but listeners did not adapt their expectations for predictive processing as expected. The model proposed by Roettger and Franke (2019) thus cannot account for the data.

The present study only examined one particular prenuclear accent: a rising pitch accent on the subject. Despite the accent being auditorily very salient (as informally judged by the authors), one could argue that this pitch accent is less prominent than the rising-falling pitch accent occurring in the EARLY NUCLEAR condition and thus does not allow a fair comparison (Baumann et al., 2015). In other words, the prominence on the subject might not be high enough to trigger attention to this part of the utterance. This is a justified hypothesis which is, however, difficult to test in German. Using a rising-falling prenuclear pitch accent results in an oddly sounding statement that strongly suggests a double contrast, i.e. both the subject and the object are discourse contrastive. A language that is more flexible with regard to the types of prenuclear pitch accents is American English. American English has been shown to exhibit rising-falling pitch accents in prenuclear positions without expressing a contrast (e.g. Chodroff & Cole, 2018; Im et al., 2018). To investigate these questions further, we conducted two additional experiments with American English listeners to test whether they can use a prenuclear rising-falling pitch accent to anticipate downstream discourse referents.[4]

4. Experiment 4 and 5: prenuclear accents in American English

4.1. Method

The method of Exps. 4 and 5 closely followed the design of Exps. 1–3. Experiments differed mainly in the

Table 5. Posterior means, 95% credible intervals, and the posterior probability of direction for relevant comparisons of all three experiments.

Parameter (omnibus model)	posterior mean	95% CI	MPE
Difference: LATE NUCLEAR (Exp. 1)–PRENUCLEAR (Exp. 2)–	−18	(−80;37)	0.74
Difference: LATE NUCLEAR (Exp. 1)–LATE NUCLEAR (Exp. 3)	−89	(−151,−25)	~1.00
Difference: PRENUCLEAR (Exp. 2)–PRENUCLEAR (Exp. 3)	−58	(−122;8)	0.95

Row 1 suggests that LATE NUCLEAR in Exp. 1 and PRENUCLEAR in Exp. 2 elicit similar anticipatory patterns, i.e. the earlier accent on the subject does not yield any anticipatory benefits; Row 2 and 3 suggest that when paired with EARLY NUCLEAR, both LATE NUCLEAR (Exp. 1) and PRENUCLEAR (Exp. 2) are quicker than if not paired with EARLY NUCLEAR (Exp. 3), suggesting that listeners selectively attend to the absence of the nuclear pitch accent on the verb if that condition is available for comparison.

generation and nature of the experimental stimuli and its consequences for the analysis. Exp. 5 differed from Exp. 4 in that it offered visual feedback to the participants.

4.1.1. Participants and procedure

31 American English listeners participated in each experiment and were recruited from the subject pool for students in introductory-level Linguistics courses at Northwestern University. All participants had self-reported normal or corrected-to-normal vision and normal hearing (Exp. 4: 18 males, 11 females, 2 non-binary; mean age = 19.6 (SD = 1), Exp. 5: 18 males, 13 females, mean age = 19.4 (SD = 1.1)).

Following protocol of Exps. 1–3, participants were seated in front of a Mac mini 2.5 GHz Intel Core i5. They controlled the experiment via a Logitech B100 corded USB Mouse. Cursor acceleration was linearised and cursor speed was slowed down. We slightly changed the narrative for this experiment. Participants were told about a *shapeshifter* called "Wuggy" which can transform into objects. There are twelve different objects that the wuggy could transform into: bagel, beaver, dollar, grizzly, ladder, lemon, lizard, mango, marble, melon, window, zebra.

On each trial, participants heard either a topic question like (9), which introduced a referent as given in the discourse, or the neutral question in (10) introducing no specific discourse content:

(9) Did the wuggy become a beaver?
(10) What does the wuggy look like?

Stimuli presentation followed the design of Exps. 1–3 (see Figure 1 above). On the response screen, the audio playback specifies which object the wuggy looks like now (see examples 11–12):

(11) The one on the screen looks like a beaver.
(12) The one on the screen looks like a mango.

Depending on the preceding question, statements in (11) and (12) are prototypically realised with different intonation contours. After a neutral question (10) offering no discourse context to relate the sentence object to, we used a default accent pattern, i.e. a high (-rising) pitch accent in the default position, i.e. on the stressed syllable of the sentence object (LEXICAL) (Figure 5).

After a polar topic question (9), the utterance in (11) can prosodically emphasise that the proposition in question is true, for example by a verum focus construction, which prosodically manifests itself here in the form of a high rising nuclear accent on the verb "looks" (henceforth EARLY NUCLEAR). Finally, the answer in (12) negates the topic question (9). It affirmatively mentions a contrastive referent, which is typically realised by a contrastive focus construction, characterised by an intonation contour with a high rising accent on the verb

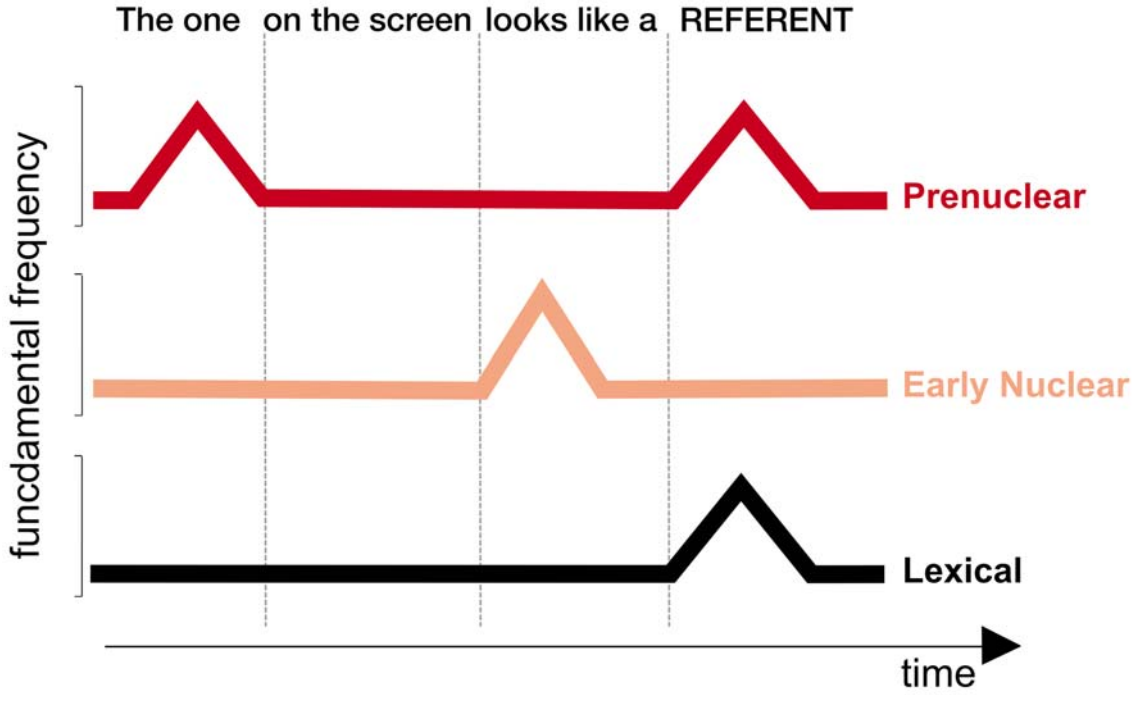

Figure 5. Schematic depiction of intonation contours used in experiments 4–5.

object. In addition to the contrastive pitch accent on the referent, the contour contains a prenuclear rising-falling pitch accent on the syntactic subject "one". We henceforth refer to this contour as the PRENUCLEAR condition. We chose this statement structure due to the following reasons: First, the accented pronoun "one" when occurring as the first accent in the utterance can be ornamental / rhythmic.[5] In the context of the experiment, a contrastive reading of the subject is unlikely (although possible), thus the prenuclear pitch accent is likely being interpreted as a cue to the discourse status of the downstream referent (if at all). Second, the distance from the prenuclear accent to both the verb and the sentence object is longer than in the German study, giving listeners ample time to integrate the early prenuclear information. All possible statements (n = 12) came with the three intonation contours (LEXICAL, EARLY NUCLEAR, and PRENUCLEAR), resulting in 36 different target sentences overall. Each target sentence was repeated four times, resulting in a total of 144 target trials (12 items * 3 conditions * 4 repetitions).

Analog to Exps. 1–3, participants saw two images and had to identify which of them is the wuggy. The response sentence ("The one on the screen looks like a X") identifies the to-be-selected referent as an "X". There were two experiments that differed only in the presence of visual feedback. Similar to Exps. 1–3, listeners in Exp. 4 only received indirect feedback when they heard the correct critical referent at the end of the utterance. Listeners in Exp. 5 additionally received visual feedback for the correct answer. When they gave their response, the eyes of the wuggy (as introduced in the instructions) were displayed on the correct referent for 500 ms. We introduced this visual feedback to ensure that listeners did not stop attending to the rest of the utterance when they gave early responses. As will be shown below, this manipulation did not have a large effect on the results.

4.1.2. Materials

Visual stimuli were taken from the BOSS corpus (Brodeur et al., 2010). Questions and statements were recorded by two trained linguists in a sound-attenuated booth at the Department of Linguistics at Northwestern University with a headset microphone (AKG C420) using 44.1 kHz/16-bit sampling. One female speaker produced the discourse-setting questions and a male speaker produced natural answers congruent with the prompting question. The male speaker produced three intonation contours: A contour with a nuclear rising-falling accent on the verb (EARLY NUCLEAR); a contour with a rising-falling accent on both the sentence subject and sentence object (PRENUCLEAR) and a contour with a rising-falling accent on the sentence object (LEXICAL). The latter contour functions as an answer to a neutral question (see example 10 above). As opposed to Exps. 1–3, we were unable to ensure that sentences across the three different intonation contours exhibit the same temporal characteristics without compromising the naturalness of the stimuli. Neither a resynthesis of the contour (e.g. Roettger & Franke, 2019) nor durational manipulations as done in Exps. 1–3 resulted in sufficiently natural sounding stimuli (as impressionistically judged by JC). We therefore refrained from durational manipulations as described for Exps. 1–3. However, to reduce at least some variability in the temporal structure of the statements and to ensure the prenuclear accent in the PRENUCLEAR condition and the nuclear accent in the EARLY NUCLEAR condition have comparable saliency, we used Praat (Boersma & Weenink, 2016) to resynthesize the f0 contours in order to achieve parity in the perceived prominence of these two accents (as impressionistically judged by JC), applying the procedure described in Appendix A2.

4.1.3. Data analysis

The mouse trajectories were processed as specified for Exps. 1–3 in §2.1.3 and turn-towards-the-target measures were calculated as specified above. Since the stimuli do not exhibit entirely comparable temporal structures, we decided to look at the turn-towards-the-target relative to the onset of the verb (see prediction below). The statistical analysis follows our specifications above: We fitted Bayesian hierarchical linear models to turn-towards-target measurements (relative to the onset of the verb) as a function of DISCOURSE context and scaled BLOCK, as well as their two-way interaction. Additionally, we included presence vs. absence of visual FEEDBACK as a fixed effect interacting with DISCOURSE and BLOCK. The models included maximal random-effect structures, allowing the predictors and their interactions to vary by participants (DISCOURSE x BLOCK) and experimental items (DISCOURSE x FEEDBACK). We used weakly informative Gaussian priors centred around zero with $\sigma = 250$ ms for all population-level regression coefficients (e.g. Gelman et al., 2008), truncated Gaussians priors for all standard deviations (mean = 0, sd = 100), and LKJ(2) priors for all correlation parameters. Four sampling chains with 2000 iterations each were run for each model, with a warm-up period of 1000 iterations.

4.1.4. Predictions

In line with Exps. 1–3, we expect that listeners can use both the presence and the absence of a nuclear pitch accent on the verb indicating that the upcoming

discourse referent is given or contrastive, respectively. Concretely, we expect compelling evidence that the difference between LEXICAL and both EARLY NUCLEAR and PRENUCLEAR is different from zero. Moreover, a rational comprehender should integrate the cue with high evidential strength (EARLY NUCLEAR = presence of early nuclear accent) substantially earlier than the cue with low evidential strength (PRENUCLEAR), predicting that responses to the EARLY NUCLEAR condition are anticipated earlier than responses to the PRENUCLEAR condition at the beginning of the experiment. The presence of the nuclear accent can be considered high in evidential strength because listeners have prior experience that this contour can map onto the givenness of the sentence object. The presence of an early prenuclear accent is assumed to be of low evidential strength because we have no reason to believe that listeners have any prior bias to interpret a pitch accent on the subject as evidence for the discourse status of the sentence object.

Regarding the integration of prenuclear cues, the following predictions can be derived. The prenuclear cue is not informative about the referent of the object at the beginning of the experiment, but a (HYPER)RATIONAL adapter should adjust its informational value when encountering sufficient evidence from the speaker. We expect that listeners will learn to exploit the presence vs. absence of the prenuclear accent to anticipate the upcoming referent. At the end of the experiment, listeners should on average turn towards the target before the onset of the verb.

4.2. Results and Discussion

Following exclusion protocol of Exps. 1–3, trials with initiation times greater than 350 ms (2.5%) and incorrect responses (0.8%) were discarded on a trial-by-trial basis. We excluded two participants (one for each experiment) due to initiation times above 350 ms in more than 15% of all trials. The remaining data were statistically analysed as specified above.

For the beginning of the experiment, there is no compelling evidence for an interaction between the presence or absence of visual feedback with either LEXICAL or the EARLY NUCLEAR condition (FEEDBACK x LEXICAL: $\beta = -27$ [−114,59]; FEEDBACK x EARLY NUCLEAR: $\beta = 58$ [−87,198]). However, there is compelling evidence that visual feedback affected listeners' anticipation behaviour for the PRENUCLEAR condition (FEEDBACK x PRENUCLEAR: $\beta = 142$ [7,269]), such that already at the beginning of the experiment, listeners anticipated the referent earlier when they had access to visual feedback. This early learning effect is thus a consequence of listeners being exposed to visual feedback in the training phase prior to the critical trials. This initial advantage does not translate, however, into any learning facilitation over the course of the experiment as indicated by highly uncertain posterior estimates for the three-way interaction of FEEDBACK x BLOCK x DISCOURSE. There is no compelling evidence that feedback affected the development of TTTs over the course of the experiment (LEXICAL: $\beta = 6$ [−4,16]; EARLY NUCLEAR: $\beta = 2$ [−20,24]; PRENUCLEAR: $\beta = -7$ [−29,14]). In the following, we therefore discuss the overall effects, collapsing the data across feedback manipulations (but see Figure 6C for a visual assessment).

Figure 6 displays the main effects (A, B) and learning effects (C) of Exps. 4 and 5. Table 6 lists numerical results. The LEXICAL condition serves us as a baseline as listeners have to wait until the acoustic information about the referential expression becomes available. The model estimates that it takes listeners 665 ms to start turning towards the target after the onset of the verb ($\beta = 665$ [635,695]). This corresponds to a time lag of 167 ms after the lexical information in the signal becomes acoustically available. As predicted, there is compelling evidence that the intonationally informed trials (EARLY NUCLEAR and PRENUCLEAR) elicit earlier TTT values than LEXICAL trials, with TTTs being earlier in the EARLY NUCLEAR condition characterised by a nuclear pitch accent on the verb ($\beta = 286$ [215,359]) than in the PRENUCLEAR condition characterised by an even earlier prenuclear pitch accent on the sentence subject ($\beta = 355$ [269,437]).

As predicted, these main effects change across the course of the experiment (see Figure 6C, Table 6). There is compelling evidence that listeners in both EARLY NUCLEAR and PRENUCLEAR trials turn towards the target faster over the course of the experiment (EARLY NUCLEAR: $\beta = -29$ [−40,−19], PRENUCLEAR: $\beta = -39$ [−48,−28]). Looking at Figure 6C, we can see that in block 1 of both experiments, EARLY NUCLEAR trials exhibit TTTs before the acoustic onset of the lexical referent (below the upper dashed line). As discussed above, even the PRENUCLEAR trials in block 1 exhibit TTTs before the acoustic onset of the referent when they get visual feedback (red line starts below the upper dashed line) and after the acoustic onset of the referent when they do not (red line starts above the upper dashed line). Most importantly, TTTs decrease rapidly during the experiment for both EARLY NUCLEAR and PRENUCLEAR trials. At the end of the experiment, the model estimates that listeners turn towards the target around 110 ms [−3,219] after the onset of the verb for EARLY NUCLEAR trials and around 124 ms [6,261] after the onset of the verb for PRENUCLEAR trials. The development throughout the experiment for the EARLY NUCLEAR and PRENUCLEAR condition is parallel, possibly suggesting that

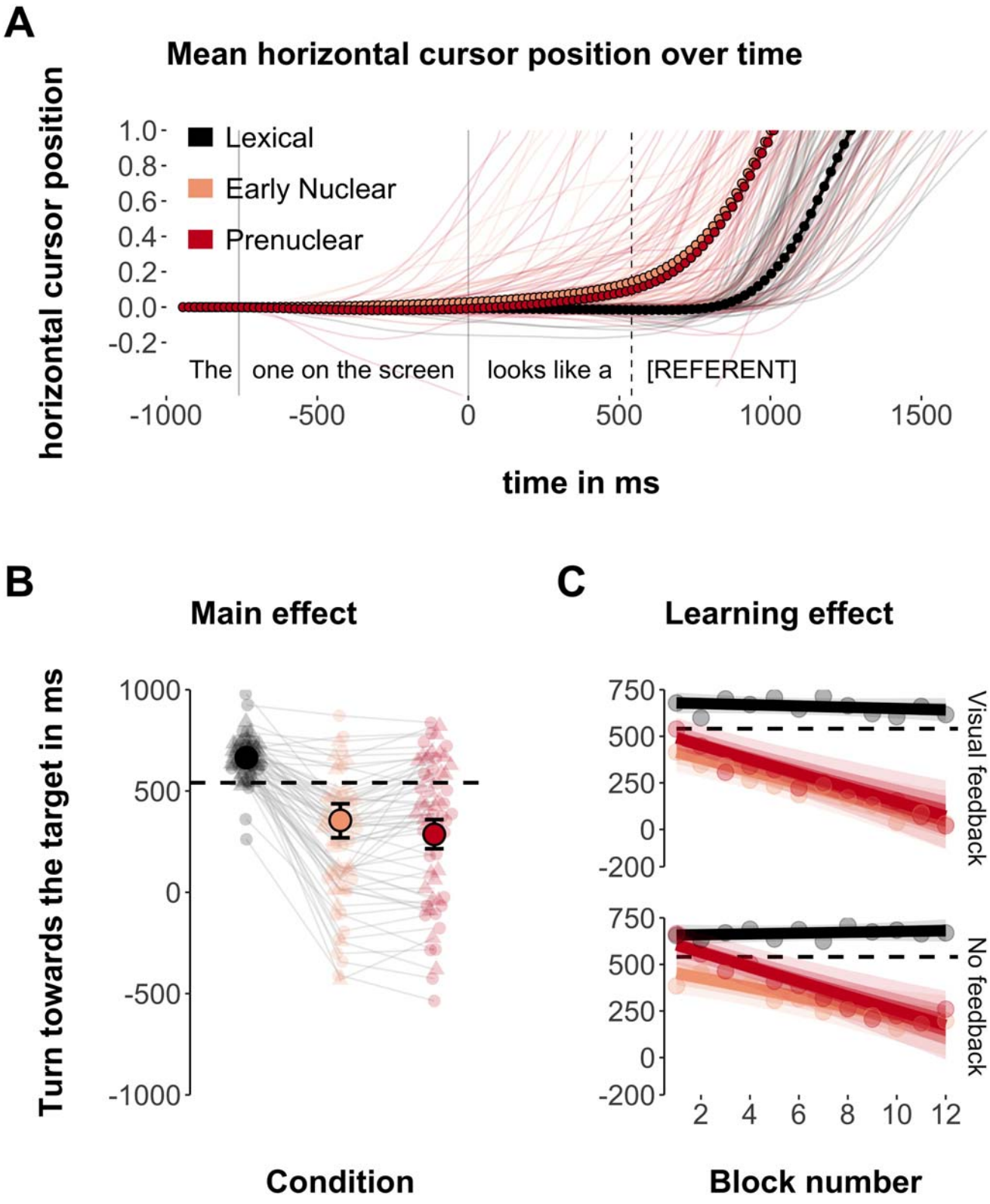

Figure 6. Collapsed results of Exps. 4 and 5. A – Horizontal cursor position of space-normalized averaged trajectories (y) plotted against time of acoustic stimulus (x) (semi-transparent lines are average values for each participant); B – Posterior means and 95% credible intervals of turn-towards-the-target measures effects across conditions (semi-transparent points are average values for each participant, circles are participants with visual feedback, triangles are participants without visual feedback); C – Posterior means and 50% / 75% / 95% credible intervals of turn-towards-the-target effects across conditions, experimental blocks and feedback condition (points indicate grand averages). Black dashed line indicates the average acoustic onset of of the sentence object.

Table 6. Posterior means, 95% credible intervals, and the posterior probability of direction for relevant comparisons of Exps. 4 and 5.

Parameter (experiments 4 and 5)	posterior mean	95% CI	MPE
Difference: LEXICAL – EARLY NUCLEAR	379	(315;379)	~1.00
Difference: LEXICAL – PRENUCLEAR	310	(239;385)	~1.00
Difference: EARLY NUCLEAR – PRENUCLEAR	−69	(−20;−112)	~1.00
Block effect: LEXICAL	−1	(−6;4)	0.59
Block effect: EARLY NUCLEAR	−29	(−40;−19)	~1.00
Block effect: PRENUCLEAR	−39	(−48;−28)	~1.00

whatever cue listeners use, they use both the presence and the absence of that cue.

The above description of the results assumes a population of listeners who behave in a uniform way. In quantitative terms, we expect turn-towards-the-target values to be normally distributed around a central tendency of the population. However, further exploratory inspections of the results might suggest otherwise. We aggregated TTTs for each participant across conditions and experimental blocks and plotted their distribution in Figure 7. Taking the LEXICAL trials as a reference point, the distribution of TTTs for both the beginning (Figure 7A) and the end of the experiment (7B) are characterised by unimodal distributions centred around the time shortly after the acoustic onset of the referent (rightmost dashed line). This pattern does not change throughout the experiment (Figure 7C, top row). Looking at the beginning of the experiment (7A), the distributions in both EARLY NUCLEAR and PRENUCLEAR trials seems to have at least one strong mode comparable to the lexical condition. This means that at the beginning of the experiment, listeners mainly waited for the lexical information to make their referential choice. Note that in the EARLY NUCLEAR condition, the distribution looks bimodal already, exhibiting another mode around the onset of the pitch accent on the verb (middle dashed line). Thus the observed variation across listeners is weakly skewed towards earlier responses, indicating a weak initial bias to integrate prosodic information already at the beginning of the experiment in the EARLY NUCLEAR condition.

Looking at the end of the experiment (7B), the distribution for both EARLY NUCLEAR and PRENUCLEAR conditions are bimodal. We can still see the late distributional peak characteristic of lexical disambiguation somewhere around 600–700 ms. However, there is another distributional peak between the onset of the prenuclear cue and the onset of the cue on the verb (between leftmost and middle dashed line) between −250 and −300 ms. This distributional peak corresponds to anticipating the referent before the acoustic onset of the verb (and the presence or absence of the pitch accent on the verb) becomes available. When looking at the development of this distribution throughout the experiment (7C, middle and bottom row), we can see the second earlier peak emerging and becoming stronger over time.

What could these patterns mean? It might mean that toward the end of the experiment in both the EARLY NUCLEAR and PRENUCLEAR condition, some listeners still wait for lexical disambiguation, as indicated by a remaining distributional peak after the referent becomes acoustically available. These listeners do not anticipate the discourse status of the referent using prosody. However, there is a group of listeners in both conditions who turn towards the target before the verb becomes acoustically available indicating the predictive use of information from the prenuclear region. These patterns suggest that the experimental data are likely generated by a mixture of (at least) two distinct anticipation patterns. In other words, an early prenuclear pitch accents *can* be used to anticipate upcoming referential expression. While some listeners *do use* these cues, other listeners *do not* use them. This interpretation has to be taken with caution, however. An inferential assessment of the assumed bimodality cannot be compellingly investigated with the present sample and warrants corroboration.

5. General Discussion

5.1. Summary

We have reported on five mouse tracking experiments to answer the question whether listeners can anticipate speakers' referential intentions based on early parts of an intonation contour. In line with previous findings, listeners were able to anticipate the discourse status of a given or a contrastive referent based on a nuclear pitch accent (e.g. Dahan et al., 2002; Ito & Speer, 2008; Kurumada et al., 2014a; Weber et al., 2006). Moreover, listeners were able to use the absence of a nuclear pitch accent, here on the verb, to anticipate an available alternative interpretation (Roettger & Franke, 2019). The latter anticipatory effect, however, is weaker than the former and manifests itself in delayed anticipatory movements. These delayed anticipatory patterns in response to the absence of an accent (on the verb) become faster over the course of the experiment, suggesting that listeners can learn to exploit this cue predictively in light of reinforcing evidence.

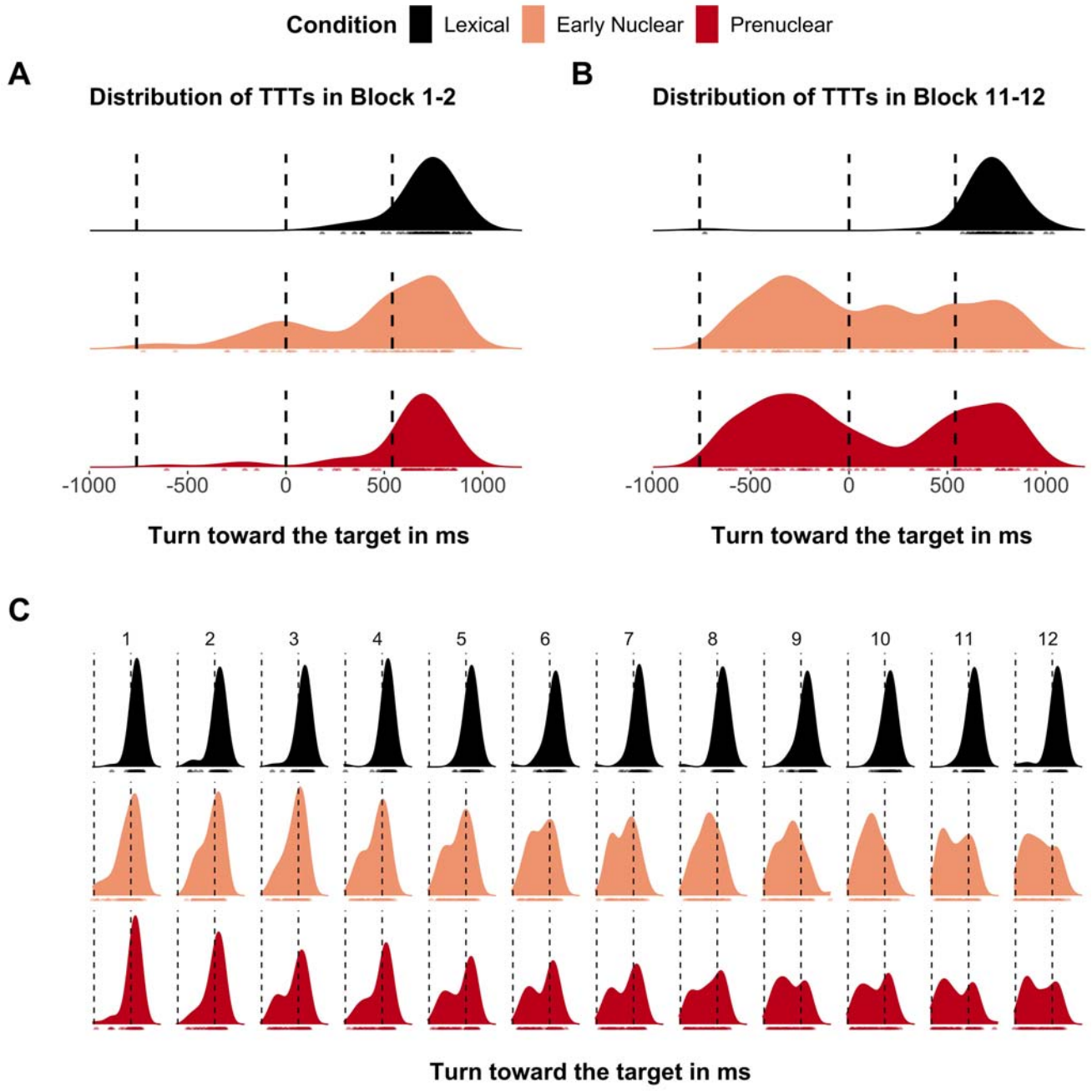

Figure 7. Kernel density plots displaying the distribution of turn toward the target values for individual listeners at the beginning of the experiment (A), the end of the experiment (B) and as developing throughout the experiments for each block separately (C) (aggregated over both Exps. 4 and 5). Dashed lines indicate the average acoustic onset of the intonational cue on the subject (pitch accent or not), on the verb (pitch accent or not) and average acoustic onset of the referent.

These findings served as a baseline for investigating the anticipatory value of prenuclear pitch accents, i.e. pitch accents that occur early in the phrase and are followed by an additional pitch accent later. Evidence from two additional experiments on German suggests that, within our experimental design, listeners do not exploit a rising prenuclear pitch accent to predict the discourse status of an upcoming referent as early as they could have. These experiments do not provide convincing evidence for a full and immediate utilisation of prenuclear information. Despite a systematic co-occurrence of prenuclear pitch accent and the discourse status of the upcoming referent within our experiment, listeners did not learn to immediately integrate this information predictively.

The prenuclear pitch accent in German was a rising pitch accent and thus a less prominent tonal event than the nuclear rising-falling pitch accents that

they were compared to. In a follow-up experiment on American English, we used a highly salient prenuclear rising-falling accent and compared it to a rising-falling nuclear pitch accent on the verb. Our data suggest that this salient prenuclear accent can, in principle, be used to anticipate the upcoming referent, however, many listeners seem to ignore the early prosodic information instead to the extent that we cannot detect a reliable immediate behavioural response to the cues. We conclude that the predictive exploitation of prenuclear pitch accent is to some extent dependent on the pitch accent type and/or its perceptual salience. This is in line with recent experimental evidence by Braun and Biezma (2019) who show that prominent prenuclear accents activate semantic alternatives, while other, less salient, pitch accents do not. Regardless of their salience, prenuclear accents are exploited for purposes of predicting downstream referential meaning in a conservative and delayed manner at best.

5.2. Positional biases in intonational processing?

Why is there only weak evidence for listeners to use information from prenuclear pitch accents to anticipate the speaker's referential intentions? We described two competing lines of research in the introduction which attribute different functional status to prenuclear accents. In order to develop a fuller understanding of the functional role of prenuclear accents, it is worth reiterating some of the arguments. In its strongest interpretation, one group of studies suggest that listeners ignore prenuclear accents and assume that listeners infer speaker intentions based on the nuclear portion of the intonation contour only. In this account, prenuclear accents are thought to be used for rhythmic purposes only (e.g. Büring, 2007; Calhoun, 2010).

In contrast, several studies have shown that speakers systematically modulate the prenuclear region of intonation contours to indicate discourse-pragmatic meaning (e.g. Braun & Asano, 2013; Petrone & D'Imperio, 2011; Petrone & Niebuhr, 2014). Listeners make also use of prenuclear pitch accents when evaluating how well intonational contours match discourse contexts (Breen et al., 2010; Birch & Clifton, 1995; Gussenhoven, 1983; Rump & Colier, 1996). The latter studies looked at whether a prenuclear pitch accent affects listeners' appropriateness judgements for broad and narrow focus, all of which are offline tasks, i.e. listeners made an evaluation after they were able to integrate the entire intonation contour. These studies suggest that if the listener has time to integrate the entire intonational information, the prenuclear accent can be informative for reference resolution.

Petrone and Niebuhr (2014) present evidence from a gating experiment on complex intonation contours. Their results suggest that the prenuclear accent contains information about the speech act meaning of the utterance which listeners pick up on. This study suggests that when not having access to the entire contour, prenuclear accents can, in principle, be used to anticipate speech act meaning.

The present study looked at the online integration of prenuclear pitch accents (see also Braun & Biezma, 2019). The present experimental evidence suggests that when listeners have access to the full intonation contour, they sometimes use prenuclear information to anticipate upcoming referential information and sometimes they don't. Whether listeners immediately integrate the prenuclear pitch accent when predicting referential intentions depends to some extent on the type of prenuclear pitch accent. A simple rise (in German) did not trigger immediate anticipation patterns of listeners, even after repeated exposure to reliable mappings between that rise and a specific referential interpretation. However, after sufficient exposure, some listeners were able to map a salient prenuclear rise-fall onto a discourse interpretation in American English.

A possible interpretation of the weak predictive nature of prenuclear accents rests on the concept of semantic integration. Heim and Alter (2006) investigated the processing of early vs. late pitch accents in the utterance using event-related potentials (ERP). Late pitch accents elicited a N400 ERP component, commonly assumed to be related to semantic processing on the sentence level (Holcomb & Neville, 1991; Kutas & Hillyard, 1984). The early pitch accents, however, elicited no late negative component, but an early positive component, identified by the authors as either a P200 or P300, both of which are associated with attentional processes. The authors speculate that the early pitch accent is heard but probably processed only in relation to the later components of the intonation contour. The idea of delayed semantic integration of prenuclear pitch accents is in line with the relative nature of prosodic prominence (Katz & Selkirk, 2011; Krahmer & Swerts, 2001; Swerts & Geluykens, 1993, 1994 see Cole, 2015, for an overview). The prosodic prominence of a word can only be judged relative to its environment including neighbouring words and surrounding intonational events.

In sum, our findings are neither in line with the strong NUCLEAR-ONLY nor in line with the strong PRENUCLEAR-MATTERS approach. It becomes clear that the

immediate integration of prenuclear accents is to some extent dependent on their form, and even highly prominent tonal events are used anticipatorily by some listeners but not by others. In cases in which listeners do not predict referential meaning based on the prenuclear accent, listeners might still process the prenuclear accent but only integrate it when the nuclear accent becomes available later in the utterance. We now turn to the question whether this behaviour is to be expected from a rational comprehender.

5.3. Rational information integration?

A rational comprehender rapidly integrates reliable information in order to probabilistically predict likely upcoming information, and adapts to changing linguistic environments (e.g. Kleinschmidt & Jaeger, 2015; Roettger & Franke, 2019). This approach successfully predicts that listeners exploit reliable mappings of intonational form and function and that the informativity of these mappings can be adjusted in light of newly encountered evidence (Kurumada et al., 2014b; Kurumada et al., 2018; Roettger & Franke, 2019). A *hyper-rational* comprehender, who flexibly uses all information at their disposal in any experimental microcosm, is furthermore expected to learn otherwise uninformative mappings if encountered systematically. In both earlier work (Roettger & Franke, 2019; Roettger & Rimland, 2020) and the present Exp. 1, listeners' anticipation patterns based on a weak cue became faster over the course of the experiment. However, in Exp. 2 and Exp. 3, listeners did not show any evidence for learning the association between an early prenuclear pitch accent and the discourse status of the sentence object, a behaviour that seems *prima facie* incompatible with a strong assumption of hyper-rationality. Moreover in Exp. 4 and 5, not all listeliners learned to predict based on the informative prenuclear pitch accent, despite our efforts to facilitate learning with a reinforcing visual cue.

A hyper-rational comprehender model seems to be at odds with these findings. However, we speculate that selectively disregarding prenuclear information might be *adaptively rational* given the limited informational value of prenuclear cues. Adaptively rational behaviour is behaviour ensuing from choice mechanisms, sensory and representational capacities that have worked well in the most frequent situations (e.g. Anderson, 1990; Chater & Oaksford, 1999, 2000; Gigerenzer & Goldstein, 1996; Hagen et al., 2012; Tversky & Kahneman, 1981). If intonational information from prenuclear positions is rarely useful for non-local disambiguation, irrespective of why that might be the case, it could actually be effort-preserving for an adaptively, resource-rational agent not to waste large processing resources on this sentential position (Lieder & Griffith, 2020). Listeners thus use their resources based on the expected utility of the cue, a proposal that has also been put forward for long-distance speech perception of segmental contrasts (Bushong & Jaeger, 2019).

The view of adaptive and resource-rational comprehender behaviour is in line with the results of a recent artificial language learning study by Kapatsinski et al. (2017). The authors presented novel intonation contours to children and adults and associated them with novel meaning categories. One contour was a M-shaped contour with two rising-falling pitch movements, an early and a late one (similar to the PRENUCLEAR contour in our Exps. 4–5). After the learning phase, learners had to map reduced versions of the contour to the learned meaning categories, with contours that either had only the early or the late rise-fall. Categorisation results suggested that adults and older children did not recognise the reduced contour with only the prenuclear accent as an instance of the full contour, but the younger children did. The authors suggest that the younger children paid attention to the holistic contour while the older children had already learned that in English there is a strong positional asymmetry in the potential for intonation contours to cue meaning. This interpretation presupposes that learners' experience with their language leads to selective attention to certain parts of the complex speech signal but not to others. A developmental trajectory from (near-)equal weighting of features to selective weighting of features has been proposed for object recognition in vision (Smith, 1989) as well as for speech sound perception (Pisoni et al., 1994). In that sense, not using the prenuclear pitch accent information could be still considered rational behaviour. Listeners rationally attend mainly to the right edge of the phrase when interpreting the referential meaning of a referent late in the sentence. In some cases, such a prior might be too strong to overcome within a short experiment (see also Kleinschmidt, 2020), but can be learned if the cue itself is highly salient and directs attention to earlier portions of the utterance.

5.4. Remaining questions and future directions

The present investigation has several limitations. Similar to studies using the visual world paradigm, our design is limited to a set of meanings that can be unambiguously illustrated by visual stimuli. Our study investigates listeners' capabilities to anticipate the information status of an upcoming referent. This is only one of many different

types of function that intonation encodes in human language and might differ with respect to the way it is processed from other functions. For example, information structural relationships such as focus are often encoded locally, i.e. primarily encoded by a tonal event on the relevant constituent. In contrast, illocutionary acts such as requesting information are commonly encoded by global modulations of the intonation contour such as pitch scaling, i.e. the entire contour including all of its low and high tones exhibit raised f0 values, and pitch excursion, i.e. the difference between low and high tones is adjusted (e.g. Haan, 2002; Hirst & Di Cristo, 1998; Ladd, 2008). Beyond non-local use of pitch, other prosodic cues such as duration have been argued to facilitate lexical access and speaker normalisation in an anticipatory way (e.g. Cutler, 1976; Brown et al., 2015). Thus, it is reasonable to assume that listeners can systematically integrate early acoustic information (including pitch) to predict upcoming information in the speech stream. One possible interpretation of our results would suggest that the absence of prediction effects for the prenuclear accent might be due to listeners' knowledge that, in German and English, these accents are not associated with non-local interpretations of information structure.

Beyond an extension towards different communicative functions, future research should further investigate how the type of acoustic event, here the pitch accent type and its perceptual salience, affects the integration of early prosodic information. Our experiments on German only examined a rising prenuclear accent, which is arguably less prominent than the rising-falling pitch accent that we used for the American English stimuli (Baumann et al., 2015). One interpretation of our results potentially speaks to how different levels of prosodic prominence do or do not direct listeners' attention to informative parts of an utterance. However, since German and English differ in how common / acceptable these different prenuclear pitch accents are, a direct comparison within either of these languages is problematic. Future studies should attempt to investigate intonational constructions that allow a fair comparison between different levels of prosodic prominence.

Finally, our exploratory analyses of the American English listeners suggest substantial variability across their predictive behaviour. Variation can potentially be attributed to two different groups of listeners, those that integrate early intonational information and those that do not. While several perception studies on intonational form-function pairs have shown vast variability across listeners (e.g. Breen et al., 2010; Cangemi et al., 2015; Roettger et al., 2019), we know very little about what it is that leads to this variability. Among the few attempts to explain interindividual variation, Bishop (2017) suggests that cognitive processing styles might account for at least some variation in intonational processing (see also Yu, 2010). Future studies should attempt to shed more light on how and why listeners differ in the way they integrate and interpret prosodic information. When doing so, we encourage these studies to consider the substantial methodological and statistical challenges that come with arguing for interindividual differences (Rouder et al., 2019).

6. Conclusion

Intonation plays a central role in human communication and can provide important early cues to speaker-intended meaning as the speech signal unfolds. However, a large body of literature suggests that not all intonational events are created equal. Production and perception studies suggest important positional asymmetries between early (prenuclear) and late (nuclear) intonational events.

Results from five mouse tracking experiments show that listeners are able to rapidly integrate some parts of an intonation contour, but not others. As demonstrated by the time at which participants started to move their mouse consistently towards a target referent, listeners picked up on the presence or absence of a nuclear pitch accent as a predictive cue to the discourse status of an upcoming referent. German listeners, however, do not use a prenuclear pitch accent as a strong predictive cue when this cue is only a rising pitch accent. When American English listeners were consistently presented with a rising-falling pitch accent, some listeners were able to update their expectations and learn to use the prenuclear cue for the anticipation of later referential intentions. However, many listeners showed no indication of such a learning effect, questioning an account to speech comprehension that is hyper-rational.

These findings suggest that intonational cue integration is constrained by certain positional asymmetries, with earlier cues being paid less attention to or weighed less heavily when updating expectations about form-function pairings. Processing theories of intonational meaning need to take these important asymmetries into account when modelling how comprehenders interpret the unfolding speech signal.

Notes

1. For our analyses, we further used the following R packages: ggbeeswarm (Clarke & Sherrill-Mix, 2017),

ggpubr (Kassambara, 2020), readbulk (Kieslich & Henninger, 2016), rstan (Stan Developer Team, 2020), rstudioapi (Ushey et al., 2020), stringr (Wickham, 2019), and tidyverse (Wickham et al., 2019).

2. Here, "heading towards the target" is operationalized by approximating the first derivative to the x- and y-coordinates of a trajectory; see function "get_TTT_derivative()" in included analysis scripts.
3. Comparisons between the model predictions and raw data indicate that the data is unlikely to be normally distributed around a single predicted mean value. Rather, it appears as if multiple distinct processes are responsible for the generation of the data. Given the nature of our primary measurement, the turn-towards-the-target (TTT), there are several instances in which listeners randomly drift toward the correct response early on (and not turning back), resulting in very early TTTs that are generated by chance rather than a genuine anticipation of the referent based on acoustic information. Given that these random drifts will occur equally often across conditions, they will not confound the comparison of groups.
4. The switch to American English was partly driven by pragmatic constraints with the first author changing institutions from the University of Cologne to Northwestern University. However, given the similarity between the German and the English intonation system, we are convinced that our findings can be informative about how listeners integrate intonational information across different sentence positions. We would like to emphasize that there are many differences between these two languages and their respective phonological systems, thus we do not consider our data as reliable evidence for cross-linguistic differences in the processing of intonation between these languages.
5. One might consider the contour in the PRENUCLEAR condition to be marked by double focus: An accented subject noun contrasts the wuggy on the screen with other wuggies seen during the experiment, while the accented object marks the referent as contrastive with regard to the topic in the preceding question. Note that this interpretation does not disqualify our claim that the accent on the subject referent is an early cue to the object referent, even if at the same time it might mark contrastive focus on "one".

Acknowledgment

Timo Roettger's work was supported by the "Zukunftskonzept" of the University of Cologne as part of the Excellence Initiative. Michael Franke's work was supported by the Priority Program XPrag.de (DFG Schwerpunktprogramm 1727). We would like to thank Nastassja Bremer and Kim Rimland for their help during data collection, as well as three anonymous reviewers, and the editor for their insightful comments and suggestions. All remaining errors are our own. Author contribution according to CRediT: **TBR**: Conceptualisation, Methodology, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualisation, Supervision, Project administration, Funding acquisition. **MF**: Conceptualisation, Methodology, Writing – Review & Editing. **JC**: Conceptualisation, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Timo B. Roettger http://orcid.org/0000-0003-1400-2739

References

Alan, C. L. (2010). Perceptual compensation is correlated with individuals' "Autistic" traits: Implications for models of sound change. *PloS one*, *5*(8), e11950. https://doi.org/10.1371/journal.pone.0011950

Ambrazaitis, G., & Niebuhr, O. (2008). *Dip and hat pattern: A phonological contrast of German?* Proceedings of the 4th international conference on Speech Prosody (pp. 269–272).

Anderson, J. R. (1990). *The adaptive character of thought*. Lawrence Erlbaum.

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*(8), 1546–1553. https://doi.org/10.1177/0956797614533705

Baumann, S., Röhr, C. T., & Grice, M. (2015). Prosodische (de-)kodierung des informationsstatus im Deutschen. *Zeitschrift Für Sprachwissenschaft*, *34*(1), 1–42. https://doi.org/10.1515/zfs-2015-0001

Bednar, J., & Page, S. (2007). Can game(s) theory explain culture?: The emergence of cultural behavior within multiple games. *Rationality and Society*, *19*(1), 65–97. https://doi.org/10.1177/1043463107075108

Bergen, L., & Goodman, N. D. (2015). The strategic use of noise in pragmatic reasoning. *Topics in Cognitive Science*, *7*(2), 336–350. https://doi.org/10.1111/tops.12144

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365–391. https://doi.org/10.1177/002383099503800403

Bishop, J. (2017). Focus projection and prenuclear accents: Evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*(2), 236–253. https://doi.org/10.1080/23273798.2016.1246745

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*. [Computer program]. Version 6.0.17.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*(3), 633–644. https://doi.org/10.2307/412039

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, *49*(4), 451–493. https://doi.org/10.1177/00238309060490040201

Braun, B., & Asano, Y. (2013). *Double contrast is signalled by prenuclear and nuclear accent types alone, not by f0-plateaux*. Proceedings of 14th annual conference of the International Speech Communication Association, (pp. 263–266).

Braun, B., & Biezma, M. (2019). Prenuclear L*+H activates alternatives for the accented word. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01993

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS One*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 306–323. http://dx.doi.org/10.1037/a0038689

Büring, D. (2007). Semantics, intonation and information structure. In G. Ramchand, & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (p. 445474). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199247455.013.0015

Bürkner, P.-C. (2016). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bushong, W., & Jaeger, T. F. (2019). *Memory maintenance of gradient speech representations is mediated by their expected utility*. Proceedings of the 41st annual conference of the Cognitive Science Society.

Buxó-Lugo, A., & Kurumada, C. (2019). *Encoding and decoding of meaning through structured variability in intonational speech prosody* [Unpublished preprint]. https://doi.org/10.31234/osf.io/9y7xj

Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, *86*(1), 1–42. https://doi.org/10.1353/lan.0.0197

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual differences in speech production and perception* (pp. 123–145). Peter Lang.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696. https://doi.org/10.1037/0278-7393.30.3.687

Chater, N., & Oaksford, M. (1999). Ten year of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. https://doi.org/10.1016/S1364-6613(98)01273-X

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese*, *122*(1), 93–131. https://doi.org/10.1023/A:1005272027245

Chodroff, E., & Cole, J. (2018). Information structure, affect and prenuclear prominence in American English. *Proceedings of 19th Annual Conference of the International Speech Communication Association*, 1848–1852. https://doi.org/10/gg466b

Clarke, E., & Sherrill-Mix, S. (2017). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. https://CRAN.R-project.org/package=ggbeeswarm

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Cole, J., Hualde, J. I., Smith, C. L., Eager, C., Mahrt, T., & Napoleão de Souza, R. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, *75*, 113–147. https://doi.org/10.1016/j.wocn.2019.05.002

Crocker, M. W. (2010). Computational psycholinguistics. In A. Clark, C. Fox, & S. Lappin (Eds.), *Computational linguistics and natural language processing* (pp. 482–513). Wiley-Blackwell.

Cruttenden, A. (1997). *Intonation*. 2nd edition. Cambridge University Press.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics*, *20*(1), 55–60. https://doi.org/10.3758/BF03198706

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452. https://doi.org/10.1002/wcs.1355

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

de Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fundamenta Mathematicae*, *17*, 298–329. https://doi.org/10.4064/fm-17-1-298-329

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

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Dotan, D., Pinheiro-Chagas, P., Al Roumi, F., & Dehaene, S. (2019). Track It to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, *23*(12), 1058–1070. https://doi.org/10.1016/j.tics.2019.10.002

Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology*, *24*(1), 2–11. https://doi.org/10.1093/beheco/ars085

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*(4), 680–703. https://doi.org/10.1016/j.wocn.2008.05.001

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language Games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Franke, M. (2009). *Signal to act: Game theory in pragmatics*. [Doctoral dissertation, Institute for Logic, Language and Computation]. University of

Amsterdam. https://dare.uva.nl/search?identifier=c530531e-9dd7-45de-a782-54eb53ed3540

Franke, M., & Jäger, G. (2016). Probabilistic pragmatics, or why bayes' rule is probably important for pragmatics. *Zeitschrift Für Sprachwissenschaft*, *35*(1), 3–44. https://doi.org/10.1515/zfs-2016-0002

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, *27*(5), 315–323. https://doi.org/10.1177/0963721417746793.

Galeazzi, P., & Franke, M. (2017). Smart representations: Rationality and evolution in a Richer environment. *Philosophy of Science*, *84*(3), 544–573. https://doi.org/10.1086/692147

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. https://doi.org/10.1016/j.visres.2010.09.027.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, M., & Baumann, S. (2007). An introduction to intonation—functions and models. In J. Trouvain, & U. Gut (Eds.), *Non-native prosody. Phonetic description and teaching practices* (pp. 25–52). de Gruyter.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107. https://doi.org/10.1016/j.wocn.2017.03.003

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

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*(2), 377–417. https://doi.org/10.1017/S0022226700007799

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Gussenhoven, C. (2011). Sentential prominence in English. In M. van Oostendorp, C. J. Ewen, E. V. Hume & K. Rice (Eds.),*The blackwell companion to phonology* (pp. 1–29). Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0116

Gussenhoven, C. (2015). Does phonological prominence exist? *Lingue e Linguaggio*, *1/2015*. https://doi.org/10/gg466f

Haan, J. (2002). *Speaking of questions*. LOT.

Hagen, E. H., Chater, N., Gallistel, C. R., Houston, A., Kacelnik, A., Kalenscher, T., Nettle, D., Oppenheimer, D., & Stephens, D. W. (2012). What can evolution do for us? In P. Hammerstein & J. R. Stevens, *Evolution and the mechanisms of decision making* (Vol. 11, pp. 97–126). MIT Press.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. Mouton de Gruyter.

Hammerstein, P., & Stevens, J. R. (2012). Six reasons for invoking evolution in decision theory. In P. Hammerstein & J. R. Stevens, *Evolution and the mechanisms of decision making* (Vol. 11, pp. 1–17). MIT Press.

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

Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, *66*, 55–68.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*(3), 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication*, *36*(1–2), 31–43. https://doi.org/10.1016/S0167-6393(01)00024-3

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*(4), 286–300.

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*(2), 217–235. https://doi.org/10.1080/23273798.2015.1083113

Im, S., Cole, J., & Baumann, S. (2018). *The probabilistic relationship between pitch accents and information status in public speech*. Proceedings of the 9th international conference on speech prosody, (pp. 508–511). https://doi.org/10/gg5mdw

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kapatsinski, V., Olejarczuk, P., & Redford, M. A. (2017). Perceptual learning of intonation contour categories in adults and 9- to 11-Year-Old children: Adults are more narrow-Minded. *Cognitive Science*, *41*(2), 383–415. doi:10/f9w62c. https://doi.org/10.1111/cogs.12345

Kassambara, A. (2020). *ggpubr: "ggplot2" Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. Discourse-new: Evidence from phonetic prominence in English. *Language*, *87*(4), 771–816. https://doi.org/10.1353/lan.2011.0076

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 736–748. https://doi.org/10.1037/0096-1523.15.4.736

Kieslich, P. J., & Henninger, F. (2016). *Readbulk: An R package for reading and combining multiple data files*. https://doi.org/10.5281/zenodo.596649

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research*

Methods, 49(5), 1–16. https://doi.org/10.3758/s13428-017-0900-z

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J., & Schulte-Mecklenbeck, M. (2019a). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kuehberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods: 2nd edition* (2nd ed.). https://doi.org/10.31234/osf.io/zuvqa

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2019b). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, *52*(1), 317–341. https://doi.org/10.3758/s13428-019-01228-y

Kleinschmidt, D. (2020). *What constrains distributional learning in adults?* PsyArXiv. https://psyarxiv.com/6yhbe/

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Knill, D. C., & Richards, W. (1996). *Perception as bayesian inference*. Cambridge University Press.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, *34*(4), 391–405. http://dx.doi.org/10.1016/S0167-6393(00)00058-3

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

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

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014b). *Rapid adaptation in online pragmatic interpretation of contrastive prosody*. Proceedings of the 36th annual meeting of the Cognitive Science Society, 36.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2018). Effects of distributional information on categorization of prosodic contours. *Psychonomic Bulletin & Review*, *25*(3), 1153–1160. https://doi.org/10.3758/s13423-017-1332-6

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, *19*, 849–856.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). Bayestestr: Describing effects and their uncertainty. *Existence and Significance Within the Bayesian Framework. Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10/gf9pds.

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

McNamara, J. M. (2013). Towards a richer evolutionary game theory. *Journal of The Royal Society Interface*, *10*(88), 20130544. https://doi.org/10.1098/rsif.2013.0544

Morett, L. M., & Fraundorf, S. H. (2019). Listeners consider alternative speaker productions in discourse comprehension and memory: Evidence from beat gesture and pitch accenting. *Memory & Cognition*, *47*(8), 1515–1530. https://doi.org/10.3758/s13421-019-00945-1

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

Petrone, C., & D'Imperio, M. (2011). From tones to tunes: Effects of the f0 prenuclear region in the perception of Neapolitan statements and questions. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 207–230). Springer Netherlands. https://doi.org/10.1007/978-94-007-0137-3_9

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, *57*(1), 108–146. https://doi.org/10.1177/0023830913495651

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121–166). The MIT Press.

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under-and over-informative prenominal adjective use. *Frontiers in Psychology*, *6*, 2035. https://doi.org/10.3389/fpsyg.2015.02035

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. https://doi.org/10.1037/a0021923

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116. https://doi.org/10.1016/j.wocn.2013.01.002

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (Un-)reliable intonation. *Cognitive Science*, *43*(7), e12745. https://doi.org/10.1111/cogs.12745

Roettger, T. B., Mahrt, T., & Cole, J. (2019). Mapping prosody onto meaning–the case of information structure in American English. *Language, Cognition and Neuroscience*, *34*(7), 841–860. https://doi.org/10.1080/23273798.2019.1587482

Roettger, T. B., & Rimland, K. (2020). Listeners' adaptation to unreliable intonation is speaker-sensitive. *Cognition*, *204*, 104372. https://doi.org/10.1016/j.cognition.2020.104372

Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why most studies of individual differences with inhibition tasks are bound to fail. Preprint at PsyArXiv. https://doi.org/10.31234/osf.io/3cjr5

Rump, H. H., & Collier, R.. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, *39*(1), 1–17. http://dx.doi.org/10.1177/002383099603900101

Rysling, A., Bishop, J., Clifton, C., & Yacovone, A. (2020). Preceding syllables are necessary for the accent advantage effect. *The Journal of the Acoustical Society of America*, *148* (3), EL285–EL288. https://doi.org/10.1121/10.0001780

Savage, L. J. (1954). *The foundations of statistics*. Dover.

Schuster, S., & Degen, J. (2020). I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, *203*, 104285. https://doi.org/10.1016/j.cognition.2020.104285

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*(1), 125–144. https://doi.org/10.1037/0033-295X.96.1.125

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/ Revue Canadienne de Psychologie Expérimentale*, *47*(2), 276–309. https://doi.org/10.1037/h0078826

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

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

Stan Development Team. (2020). *RStan: The R interface to Stan*. http://mc-stan.org/

Swerts, M., & Geluykens, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica*, *50*(3), 189–196. https://doi.org/10.1159/000261939

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37* (1), 21–43. https://doi.org/10.1177/002383099403700102

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

Tomlinson, J. M., Gotzner, N., & Bott, L. (2017). Intonation and pragmatic enrichment: How intonation constrains ad hoc scalar inferences. *Language and Speech*, *60*(2), 200–223. https://doi.org/10.1177/0023830917716101

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Ushey, K., Allaire, J. J., Wickham, H., & Ritchie, G. (2020). *rstudioapi: Safely Access the RStudio API*. https://CRAN.R-project.org/package=rstudioapi

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton university press.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392. https://doi.org/10.1177/00238309060490030301

Wickham, H. (2019). *stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686