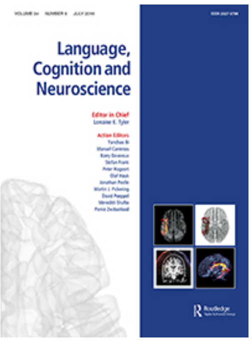

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Listener-oriented phonetic reduction and theory of mind

Rory Turnbull

LANGUAGE, COGNITION AND NEUROSCIENCE
2019, VOL. 34, NO. 6, 747–768
https://doi.org/10.1080/23273798.2019.1579349

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Listener-oriented phonetic reduction and theory of mind

Rory Turnbull[a,b]

[a]Department of Linguistics, University of Hawai'i at Mānoa, Honolulu, HI, USA; [b]Department of Linguistics, Ohio State University, Columbus, OH, USA

ABSTRACT
Predictable words tend to be phonetically reduced relative to unpredictable words. Under "listener-oriented" accounts of this phenomenon, the talker has tacit knowledge of their interlocutor's mental state. These theories consequently predict that individual variation in theory of mind is related to magnitude of probabilistic phonetic reduction. The current study tests this prediction for three acoustic variables (word duration, vowel duration, and vowel dispersion) in two definitions of predictability (contextual predictability and discourse mention). A relationship between individual variation in theory of mind and phonetic reduction was observed only for semantic predictability, and in the direction opposite to that predicted by listener-oriented theories. Taken together, these results are not consistent with the predictions of a strong interpretation of listener-orientation in speech production.

ARTICLE HISTORY
Received 2 May 2017
Accepted 3 December 2018

KEYWORDS
Listener modelling; discourse mention; semantic predictability; vowel dispersion; duration

1. Introduction

Phonetic reduction is pervasive in natural speech. Sounds and syllables are commonly shorter, less prominent, or even deleted entirely. A well-established finding in the literature is that the likelihood or magnitude of reduction is inversely correlated with various measures of word predictability. For example, the second mention of a word in a discourse generally has a shorter word duration, a shorter vowel duration, and is less intelligible in isolation than the word's first mention (Baker & Bradlow, 2009; Bard & Anderson, 1994; Bard et al., 2000; Bard, Lowe, & Altmann, 1989; Burdin & Clopper, 2015; Fowler, 1988; Fowler & Housum, 1987; Fowler, Levy, & Brown, 1997; Galati & Brennan, 2010; Hawkins & Warren, 1994; Kahn & Arnold, 2012, 2015; Kaiser, Li, & Holsinger, 2011; Lam & Watson, 2010, 2014; Pate & Goldwater, 2011; Sasisekaran & Munson, 2012; Shields & Balota, 1991; Turnbull, 2017b; Vajrabhaya & Kapatsinski, 2011). Likewise, words which are predictable given the preceding context are less intelligible in isolation than contextually unpredictable words (Lieberman, 1963). Acoustically, this distinction has been observed to manifest itself in shorter word and vowel durations (Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Clopper & Pierrehumbert, 2008; Engelhardt & Ferreira, 2014; Ernestus, Hanique, & Verboom, 2015; Gahl & Garnsey, 2004; Hunnicutt, 1985, 1987; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Moore-Cantwell, 2013; Pate & Goldwater, 2011; Pluymaekers, Ernestus, & Baayen, 2005; Tily & Kuperman, 2012; Turnbull, 2015b), less disperse vowels (Aylett & Turk, 2006; Clopper & Pierrehumbert, 2008; Jurafsky et al., 2001), less prosodic prominence (Kaland, Krahmer, & Swerts, 2014; Turnbull, 2017b; Turnbull, Burdin, Clopper, & Tonhauser, 2015; Watson, Arnold, & Tanenhaus, 2008), more nasal place assimilation (Turnbull, Seyfarth, Hume, & Jaeger, 2018), and more phoneme deletion (Cohen Priva, 2015; Tily & Kuperman, 2012; Turnbull, 2018) in predictable words relative to unpredictable words. While it is clear that there are many factors relating to predictability which lead to phonetic reduction (see Clopper & Turnbull, 2018, for review), the causal mechanisms underlying this phenomenon are less clear.

1.1. "Listener orientated" accounts of predictability-based reduction

Several proposals explaining this phenomenon have been suggested. Several of these proposals can be grouped as "listener-oriented" accounts (see Turnbull, 2015a, Chapter 1, for review). These accounts (e.g. Aylett, 2000; Aylett & Turk, 2004, 2006; Frank & Jaeger, 2008; Galati & Brennan, 2010; Genzel & Charniak, 2003; Jaeger, 2013; Jaeger & Tily, 2011; Pate & Goldwater, 2015; Qian & Jaeger, 2012; Ramscar & Baayen, 2013; Schober, 1993; Turk, 2010; Van Son & Pols, 2003; Van Son & Van Santen, 2005, *inter alia*) hold that there are

CONTACT Rory Turnbull rory.turnbull@hawaii.edu
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2019.1579349

direct and active processes that serve to enhance the perceptibility of speech in low-predictability conditions. The inhibition of these processes, the accounts further posit, is the cause of phonetic reduction in speech. A classic formulation of this view is Björn Lindblom's "Hyper- and Hypo-Articulation Theory" (H&H theory), where the talker is assumed to have "tacit awareness of the listener's access to sources of information independent of the signal and his [the listener's] judgement of the short-term demands for explicit signal information" (Lindblom, 1990a, p. 403). In other words, the talker knows which aspects of the signal are highly predictable due to context or language structure; these forms are therefore somewhat redundant and the talker can reduce them (i.e. "hypoarticulate"). The talker also knows which aspects of the signal the listener may misperceive; the talker can emphasise these forms (i.e. "hyperarticulate"). The whole act of speech production is a balance between conservation of effort (causing hypoarticulation) and attending to listener need (causing hyperarticulation).

Some proposals, such as those of Pierrehumbert (2002, 2003), have a key role for the listener in the origin and propagation of predictability effects. However, these accounts are not listener-oriented in the strict sense as they do not require the talker to have a mental model of the listener. Rather, these proposals have been termed "passive" or "evolutionary" accounts (Clopper & Turnbull, 2018; Turnbull, 2015a), as they do not posit active cognitive mechanisms to account for predictability effects.[1] The term "listener-oriented" is only applied to theories for which the talker is in some way taking into account the perspective of the listener. This definition therefore includes theories like H&H theory which state that predictability effects exist as a means to ensure smooth and uninterrupted communication between interlocutors; that is, speech is an adaptive and goal-oriented process. Such proposals include the Smooth Signal Redundancy Hypothesis (Aylett, 2000; Aylett & Turk, 2004, 2006; Turk, 2010), and those of Flemming (2010), Fox Tree and Clark (1997), Levy and Jaeger (2007), and Rosa, Finch, Bergeson, and Arnold (2015).

A strong interpretation of the listener-oriented account was made explicit by Lindblom (1990a), who referred to the talker's "tacit awareness of the listener's access to sources of information independent of the signal" (p. 403). Talkers are "mind readers", who have the ability to "imagine the world from another person's point of view" (Lindblom, 1990b, p. 228). Other scholars have had similar readings of these claims: the talker "has the listener in mind" (Moon, 1995, pp. 488–489), possesses "an up-to-date model of the listener's current knowledge" (Bard & Aylett, 2005, p. 176), and makes "an effort to use knowledge about the listener to anticipate likely errors or messages" (Pate & Goldwater, 2015, p. 4) (see also Donnarumma, Dindo, Iodice, & Pezzulo, 2017 and Oesch & Dunbar, 2017). This interpretation of the listener-oriented account holds that the talker must be able to create, maintain, and update a detailed mental representation of their interlocutor's knowledge, beliefs, intentions, desires, and emotions in real time – that is, the talker must possess a well-developed theory of mind, the ability to impute complex mental states to others. From this interpretation it follows that individual variation in theory of mind ability is linked to the extent and application of predictability-based reduction.[2] This paper tests this prediction by examining correlations between scores in theory of mind tasks and the magnitude of acoustic reduction in speech production.

1.2. Theory of mind and listener orientation

Theory of mind (ToM) is the ability to attribute mental states to others; if someone possesses a ToM, it means "that the individual imputes mental states to others" (Premack & Woodruff, 1978, p. 515). This definition naturally includes a wide range of potential states, including those of knowledge, intention, purpose, desire, emotion, and so on, for an individual to attribute to others. For present purposes, we will use Premack & Woodruff's above definition, necessarily excluding some definitions which would have "empathy" rather than ToM be responsible for emotional mental states (see Decety & Jackson, 2004).[3] ToM as thus defined is not directly observable and must be inferred from behavioural measures.

It follows from a strong interpretation of the listener-oriented account that the talker must employ a variety of ToM skills in order to be able to communicate efficiently (i.e. to reduce and enhance in the appropriate places). According to this reasoning, without full command of these ToM abilities, the talker is not able to know when to reduce and when to enhance. Regardless of one's theory of how the speech production process unfolds, the listener-oriented account requires that the talker's knowledge of the listener exerts some influence upon speech. Figure 1 shows a schematised and simplified boxology of this process. The speech production process has some higher level processes (such as syntax, phonological planning, and other details) which feed into some process of phonetic planning. From the phonetic planning, speech is produced. The figure is deliberately agnostic as to whether this "phonetic planning" is in terms of symbolic phonological units, articulatory kinematics, gestural scoring, or some other process.

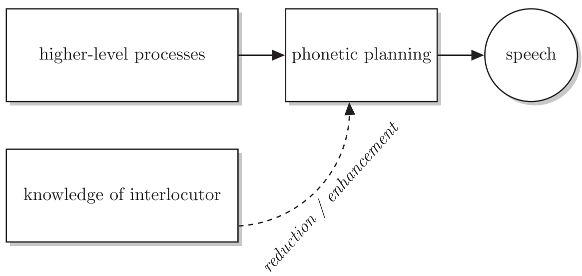

Figure 1. Simplified boxology of influence of knowledge of one's interlocutor on speech production. See text for details.

The precise details are irrelevant. In any case, according to a listener-oriented account, the knowledge of one's interlocutor intervenes at some point to reduce or enhance certain phrases, words, syllables, phones, or some other unit of planning. If it is known that the interlocutor is likely to find some unit predictable, then that unit can be reduced. Unpredictable units can be enhanced. Thus this subsystem bears influence on the speech production process. In an extreme case, if there is no knowledge of the interlocutor, that is, the talker does not possess a theory of mind, then there is no influence and thus no observed predictability effects. However, all neurotypical individuals possess some degree of ToM; the variation is in how easily they are able to use it. Variation in use of ToM among the neurotypical population follows from the notion that taking into account the perspective of others is an effortful process which takes time and energy (e.g. Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003; Lin, Keysar, & Epley, 2010). In the case of the present boxology, then, talkers with a weaker ToM will be delayed in their calculations of interlocutor knowledge and its relevance for speech communication, relative to talkers with a stronger ToM. This delay entails the subsystem having less time to exert its influence on phonetic planning before the talker moves on to planning the next word. As a consequence, a smaller degree of enhancement or reduction predicted. For talkers with a stronger ToM, the calculations of what to enhance and reduce will be that much faster and easier, and thus the effects are predicted to be larger.

This paper presents the results of two experiments into the role of ToM in phonetic reduction. Reduction arising due to semantic predictability and second mention were examined. To preface our results, individual variation in ToM was not found to systematically correlate with phonetic reduction due to second mention. A correlation between theory of mind and reduction due to semantic predictability was observed, such that talkers with a stronger ToM exhibited smaller reduction effects than talkers with a weaker ToM. This effect is in the opposite direction to that predicted by the strong listener-oriented account. Taken together, the results of this study are not consistent with a strong listener-oriented account of predictability effects.

2. Experiment 1: semantic predictability-based reduction and second mention reduction

2.1. Introduction

The first experiment investigated correlations between individual differences in ToM and phonetic reduction due to semantic predictability and discourse mention (i.e. second mention reduction).

2.2. Method

Participants completed three tasks designed to assess their ToM ability and two speech production tasks. The experiment was administered via computer in a double-walled sound attenuated booth.

2.2.1. Participants

Twenty-one eligible participants completed the task for partial course credit. Eligibility was restricted to native monolingual speakers of American English with no history of speech, language, or hearing disorders. None of the participants reported any history of autism-spectrum conditions.

2.2.2. Speech production tasks

In the first task, the participants were instructed to read sentences aloud. The stimuli consisted of 41 matched high- and low-predictability sentence pairs, drawn from the SPIN sentences (Kalikow, Stevens, & Elliott, 1977), which varied in the semantic predictability of the final (target) word. An example pair is shown in sec(1) and sec(2); the full list of sentences used is provided in Appendix 1. This task was intended to elicit phonetic reduction on the target word in the high-predictability context.

(1) For your birthday I baked a cake. (High predictability)
(2) Tom wants to know about the cake. (Low predictability)

The sentences were blocked by predictability condition and presented in random order. Block order was counterbalanced between participants. Each sentence was presented individually on a computer screen for 3500 ms, before advancing to the next sentence.

In the second task, participants read aloud five narrative paragraphs. These paragraphs were adapted from those used by Clopper, Mitsch, and Tamati (2017)[4], and featured repetitions of target items in different contexts to elicit second mention reduction. The paragraphs are provided in full in Appendix 2. The paragraphs had a total of 40 target words, each said twice, for a total of 80 target word tokens. Each paragraph was presented on screen, and trial advancement was self-paced.

Recordings were made via a Shure SM10A head-mounted microphone connected directly to a Marantz PMD661 digital recorder.

2.2.3. Theory of mind assessment

Following the reading tasks, the participants completed three ToM assessment tasks: the Reading the Mind in the Eyes (RMITE) test, the autism-spectrum quotient (AQ) questionnaire, and the Strange Stories (SS) task.

In the RMITE test, the participant is presented with a photograph of the eye region of the face, and asked to choose one of four adjectives to best describe what the person in the photograph is feeling (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, & Jolliffe, 1997). This test assesses ability to infer complex emotional states from simple pictures, and thus serves as an indirect measure of theory of mind. The test has been shown to be a reliable and stable measure of ToM in non-clinical adult populations (Fernández-Abascal, Cabello, Fernández-Berrocal, & Baron-Cohen, 2013; Vellante et al., 2013).

The AQ questionnaire was designed to measure autistic traits in adults of normal intelligence (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Autistic traits are relevant to ToM insofar as the lack of theory of mind has been proposed as one of the primary cognitive deficits observed in people with autism spectrum disorders (Baron-Cohen, Leslie, & Frith, 1985). The questionnaire takes the form of thirty statements such as “I am good at social chit-chat” to which the test-taker responds on a four-point agree/disagree scale.[5] Due to its ease of administration and scoring, a high test–retest reliability, and a sensitivity to some aspects of ToM which other tasks lack (Senju, Southgate, White, & Frith, 2009; Stevenson & Hart, 2017; Stewart, Allison, Baron-Cohen, & Watson, 2015), the AQ has seen wide use in developmental research since its introduction (but cf. Bishop & Seltzer, 2012; Lundqvist & Lindner, 2017; Nishiyama et al., 2014). It has also been used to aid in the investigation in the role of autistic traits in speech processing (Grice, Krüger, & Vogeley, 2016; Stewart & Ota, 2008; Yu, 2010) and speech production (Mielke, Nielsen, & Magloughlin, 2013; Ward, 2013; Yu, 2016; Yu, Abrego-Collier, & Sonderegger, 2013; Yu, Grove, Martinović, & Sonderegger, 2011).

Finally, in the Strange Stories (SS) task, the participant reads short, paragraph-length stories and answers a free response question about the story. Target questions focus on the participant’s understanding of the motivations and relationships between the story’s characters (Happé, 1994; Jolliffe & Baron-Cohen, 1999; White, Hill, Happé, & Frith, 2009). Responses are scored based on whether or not the answer made reference to the mental state of others. A higher score therefore indicates a greater degree of ToM ability – we refer to this score as the “raw SS” score. The task also includes control questions which examine the participant’s ability to reason about physical states relating to humans, non-human physical states, and simple text comprehension. Subtracting the scores on the target questions from the scores on the control questions yields a normalised SS score. A normalised SS score near zero indicates that the participant performed as well on the target questions as on the comprehension questions, while a positive score indicates that the participant performed poorly on the mental state questions relative to the comprehension questions, and likely has a weaker ToM ability.

The published version of the SS task is written in British English, using syntax and lexis that may be confusing or distracting for an American English-speaking participant population. Consequently, each of the stories and questions were reworded into a format that American English speakers found more natural. This rephrasing was performed in consultation with a team of six native speakers of American English and one native speaker of British English. Full story and question sets are provided in Appendix 3. A total of twenty-eight stories (eight target stories) were presented in random order.

For ease of reference, the directionality of the scores derived from each of these tasks is summarised in Table 1.

2.2.4. Acoustic measurements

Measures of word duration in milliseconds, vowel duration in milliseconds, and vowel midpoint F1 and F2 in ERB[6] were taken of the target words of both production tasks. Vowel dispersion was calculated following the method outlined by Turnbull (2017a): the Euclidean

Table 1. Summary table of how to interpret directionality of various ToM measures.

ToM measure	Larger value means…	Smaller value means…
AQ	Poorer ToM	Stronger ToM
RMITE	Stronger ToM	Poorer ToM
Raw SS	Stronger ToM	Poorer ToM
Normalised SS	Poorer ToM	Stronger ToM

distance from each vowel token to the grand mean of by-vowel means for that talker. Segmentation was carried out by both the author and undergraduate research assistants trained in phonetic segmentation. This process was not done blindly – the high- or low-predictability status of each token was known at the time of segmentation. This non-blind segmentation introduces the possibility of bias – conscious or subconscious – influencing the duration measurements. Note that such a bias would not be able to influence vowel formant measures (which were estimated at the midpoint of the vowel), nor would it influence the critical hypothesis of this study (as the ToM scores of the participants was not known at the time of segmentation).

Many of the target words in the paragraph-reading task were sentence-medial and may or may not have been followed by a prosodic break. To control for potential effects of phrase-final lengthening, the presence or absence of a pause after every target word was also coded.

2.2.5. Analysis

For the sentence-reading task, out of a total of 1722 target tokens (21 participants × 82 target tokens), 72 disfluent or restarted tokens were excluded, resulting in 1650 tokens being entered into the analysis. For the paragraph-reading task, out of a total of 1680 target tokens (21 participants × 80 target tokens), 63 tokens were excluded, resulting in 1617 tokens being entered into the analysis.

Linear mixed effects regression models were constructed for each production task for each of the acoustic variables – word duration, vowel duration, and vowel dispersion. Due to potential problems in model interpretation due to collinearity among the ToM measures, a separate model for each ToM task was constructed. Fixed effects were predictability condition or discourse mention (high predictability versus low predictability, with high as reference level; or first mention versus second mention, with first as reference level), ToM task score, and a 2-way interaction between predictability/mention and the ToM measure. The models with discourse mention additionally had a fixed effect of whether the target word was produced before a pause or not. All continuous variables were centred around the mean before being entered into the model. Random intercepts for talker and word identity were included, with random slopes for predictability/mention condition for both talker and word. In all models, *p*-values were calculated by treating the absolute *t*-statistic as if it came from a *t*-distribution with degrees of freedom equal to the number of data points minus the number of model parameters (Baayen, 2008).

To assess the possibility of outlier participants or items unduly affecting the analyses, influence diagnostics were calculated for each model (Loy & Hofmann, 2014). These diagnostics included measures of influence on parameter estimates (Cook's distance and MDFFITS), influence on the precision of parameter estimates (CovTrace and CovRatio), and influence on variance estimates (relative variance change). For each model, these diagnostics were calculated for each participant and for each target word. In the results section, diagnostics are only reported for unusually extreme values (Cook & Weisberg, 1982).

For all models presented here, fixed and random effect correlation tables are provided in Appendix 4.

2.2.6. Predictions

If the strong interpretation of the listener-oriented account is correct, then it is expected that participants with better theory of mind abilities ought to produce larger and more consistent differences between high- and low-predictability items and between first and second mentions than participants with poorer theory of mind. In other words, this interpretation predicts a significant interaction between predictability/mention and the ToM measures.

2.3. Results

2.3.1. Individual difference scores

Variation in the ToM scores was within the normal range established in prior work. AQ scores ranged from 90 to 128, with a mean of 111.48 ($SD = 11.02$), similar to Austin's (2005) reported mean AQs of 109.4 ($SD = 14$) for a sample of male college students and 104.1 ($SD = 13.5$) for a sample of female college students. RMITE scores ranged from 15 to 33, with a mean of 26.92 ($SD = 4.77$), close to Baron-Cohen, Wheelwright, Hill et al.'s (2001) reported neurotypical mean of 26.2 ($SD = 3.6$). Raw SS scores ranged from 16 to 53, with a mean of 42.62 ($SD = 8.23$); and normalised SS scores ranged from -1 to 7, with a mean of 3.10 ($SD = 2.23$). There are no published expected population means for the Strange Stories score for comparison. Table 2 depicts a

Table 2. Correlation matrix of the individual difference scores in experiment 1, also including *p*-values for the significance of each pairwise comparison.

	AQ score		RMITE		SS (raw)	
	r	(*p*)	*r*	(*p*)	*r*	(*p*)
RMITE score	−.154	(.505)				
Strange stories (raw)	−.383	(.087)	.673	(.001)		
Strange stories (normalised)	.129	(.578)	.163	(.480)	−.240	(.295)

correlation matrix between all four of these individual difference scores, showing both Pearson's r and p-values. As can be seen, the RMITE and the raw SS scores were significantly positively correlated.[7]

2.3.2. Semantic predictability-based reduction results

The outputs of the regression analyses predicting word duration in the sentence-reading task are summarised in Table 3. In all models, a simple effect of predictability condition was observed, such that unpredictable words (M = 429 ms) were longer than predictable words (M = 410 ms). This result is consistent with prior research (Lieberman, 1963; Turnbull & Clopper, 2013), and demonstrates that phonetic reduction occurred. No simple effects of the individual ToM scores (AQ, RMITE, raw SS, and normalised SS) were observed, but significant interactions were observed. Predictability significantly interacted with AQ score, RMITE score, and raw SS score. These interactions were such that participants with higher AQ scores had a larger difference in word duration between the predictable and unpredictable contexts (i.e. a larger extent of phonetic reduction) than participants with lower AQ scores; participants with lower RMITE scores or lower raw SS scores had larger differences between predictable and unpredictable contexts than participants with higher RMITE scores or higher raw SS scores. Bearing in mind that high AQ, low RMITE, and low raw SS score all indicate a poorer ToM (cf. Table 1), a clear pattern emerges: participants with less ToM ability have larger or more extensive reduction in word duration between the low and high predictability conditions. These interaction between predictability and ToM scores are visualised in Figure 2. As the graph shows, one participant had an unusually low Strange Stories score (more than three standard deviations from sample mean) and appears to have a disproportionate influence on the result. Although this participant's Cook's distance was not overly high (0.348), removal of this participant led to the Predictability by SS score interaction no longer being significant (p = .180).

The outputs of the regression analyses predicting vowel duration in the sentence-reading task are summarised in Table 4. In all models, a simple effect of predictability condition was observed, such that vowels in unpredictable words (M = 188 ms) were longer than vowels in predictable words (M = 181 ms). A single significant interaction of predictability with raw SS score was observed, such that participants with lower SS scores exhibited greater phonetic reduction than participants with higher SS scores. This effect is depicted in Figure 3, and appears to be driven by a single participant who had an exceptionally low score on the Strange Stories task. This participant's Cook's distance was unusually high for the raw SS and RMITE models (1.286 and 2.167, respectively), indicating an oversized influence on the overall results. After removing this participant, the predictability and SS interaction was no longer significant (p = .798); the predictability and RMITE interaction, formerly insignificant, became significant (β = 1.451, t = 2.041, p = .041). This interaction is in the opposite direction to that observed in the word duration model. These effects are shown in the red trendline of Figure 3.

Table 5 summarises the output of the regression analysis predicting vowel dispersion in the sentence-reading task. In all models, a significant simple effect of predictability was observed, such that vowels in unpredictable words (M = 10.471 ERB) were more disperse than vowels in predictable words (M = 10.420 ERB). A significant interaction between predictability and AQ was also observed, such that participants with higher AQ scores exhibited greater phonetic reduction than participants with lower AQ scores.

Taken together, these results suggest that individuals with poorer theory of mind (higher AQ scores, lower RMITE scores) in fact exhibit *larger* predictability effects than individuals with more theory of mind. This conclusion is unexpected: the listener-oriented account predicts that poorer theory of mind should lead to *less* listener adaptation, not more. This result suggests that our current understanding of the AQ and its relationship to ToM and listener modelling is inadequate and deserves further investigation.

Table 3. Model outputs for word duration models in the sentence production task of Experiment 1.

	β	SE	t	p
Intercept	−8.372	18.343	−0.456	0.648
Pred.: Low	20.151	6.091	3.308	0.001
AQ score	−0.060	1.021	−0.059	0.953
Pred:AQ	1.284	0.581	2.207	0.027
	β	SE	t	p
Intercept	−8.160	18.215	−0.448	0.654
Pred.: Low	19.969	5.712	3.496	0.000
RMITE score	1.905	2.115	0.901	0.368
Pred:RMITE	−3.327	1.155	−2.881	0.004
	β	SE	t	p
Intercept	−8.264	18.296	−0.452	0.652
Pred.: Low	20.021	5.350	3.742	0.000
SS (raw)	0.685	1.302	0.526	0.599
Pred:SS(r)	−2.292	0.656	−3.494	0.000
	β	SE	t	p
Intercept	−8.368	18.342	−0.456	0.648
Pred.: Low	20.252	6.781	2.987	0.003
SS (normalised)	0.303	4.835	0.063	0.950
Pred:SS(n)	−1.299	3.075	−0.422	0.673

2.3.3. Second mention reduction results

The model outputs from the word duration, vowel duration, and vowel dispersion models from the

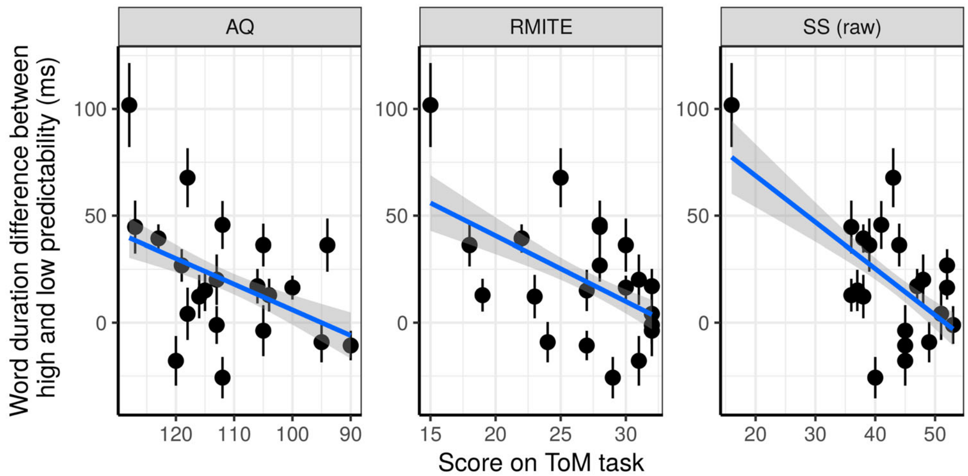

Figure 2. Reduction in word duration from low to high predictability conditions as a function of talker AQ score, RMITE score, and raw Strange Stories score. A positive duration difference means the low-predictability production was longer than the high-predictability production. Note that the axis of the AQ task is flipped for consistent interpretation of ToM skill – participants with poorer ToM skills cluster towards the left, while those with better ToM skills cluster towards the right. Graph depicts subject means and standard errors with linear trend and confidence interval overlaid.

paragraph-reading task are shown in Tables 6, 7, and 8, respectively. As can be seen, significant effects of mention were observed for all models of all three acoustic variables: second mentions were reduced (M_{word} = 313 ms, M_{vowel} = 115 ms, M = 2.778 ERB) relative to first mentions (M_{word} = 337 ms, M_{vowel} = 126 ms, M = 2.884 ERB). These effects are consistent with prior literature on second mention reduction (Baker & Bradlow, 2009; Burdin & Clopper, 2015). Additionally, consistent with the effects of phrase-final lengthening (Turk & Shattuck-Hufnagel, 2000, 2007), words and vowels preceding a pause were significantly longer (M_{word} = 378 ms, M_{vowel} = 143 ms) than words not preceding a pause (M_{word} = 315 ms, M_{vowel} = 116 ms). Significant interactions with ToM variables were observed on vowel dispersion for raw SS scores and RMITE scores, such that the difference between first and second mentions was larger for participants with stronger ToM than for participants with poorer ToM, suggesting a greater degree of reduction in the second mentions. As with the vowel duration model, this significant interaction disappeared when the participant with an outlier SS score was dropped from the model (p = .212), although the interaction with RMITE remained ($\beta = -0.012$, $t = -2.188$, p = .034).

Table 4. Model outputs for vowel duration models in the sentence production task of Experiment 1.

	β	SE	*t*	*p*
Intercept	−4.195	9.916	−0.423	0.672
Pred.: Low	8.804	4.217	2.088	0.037
AQ score	−0.360	0.833	−0.433	0.665
Pred:AQ	0.517	0.407	1.269	0.205
	β	SE	*t*	*p*
Intercept	−4.307	9.920	−0.434	0.664
Pred.: Low	8.810	4.361	2.020	0.044
RMITE score	−0.706	1.764	−0.400	0.689
Pred:RMITE	−0.461	0.894	−0.515	0.606
	β	SE	*t*	*p*
Intercept	−4.318	9.859	−0.438	0.661
Pred.: Low	8.742	3.840	2.276	0.023
SS (raw)	−0.727	1.062	−0.685	0.494
Pred:SS(r)	−1.164	0.479	−2.431	0.015
	β	SE	*t*	*p*
Intercept	−4.099	9.788	−0.419	0.675
Pred.: Low	8.870	4.372	2.029	0.043
SS (normalised)	3.583	3.876	0.924	0.355
Pred:SS(n)	0.750	2.001	0.375	0.708

2.4. Discussion

2.4.1. Phonetic reduction and theory of mind

This experiment tested interactions between two kinds of phonetic reduction – semantic predictability-based reduction and second mention reduction – and individual variation in theory of mind. Overall, reduction was observed, with predictable words being reduced relative to unpredictable words, and second mentions being reduced relative to first mentions. As outlined in the introduction, a strong interpretation of the listener-oriented theory of predictability-based phonetic reduction holds that individuals with less adept ToM skills should produce less phonetic reduction than individuals with more adept ToM skills. Two interactions were observed in a direction consistent with this hypothesis: reduction in vowel duration for predictably words was larger for participants with better ToM as measured by the RMITE score; and reduction in vowel dispersion for second mentions was larger for participants with better

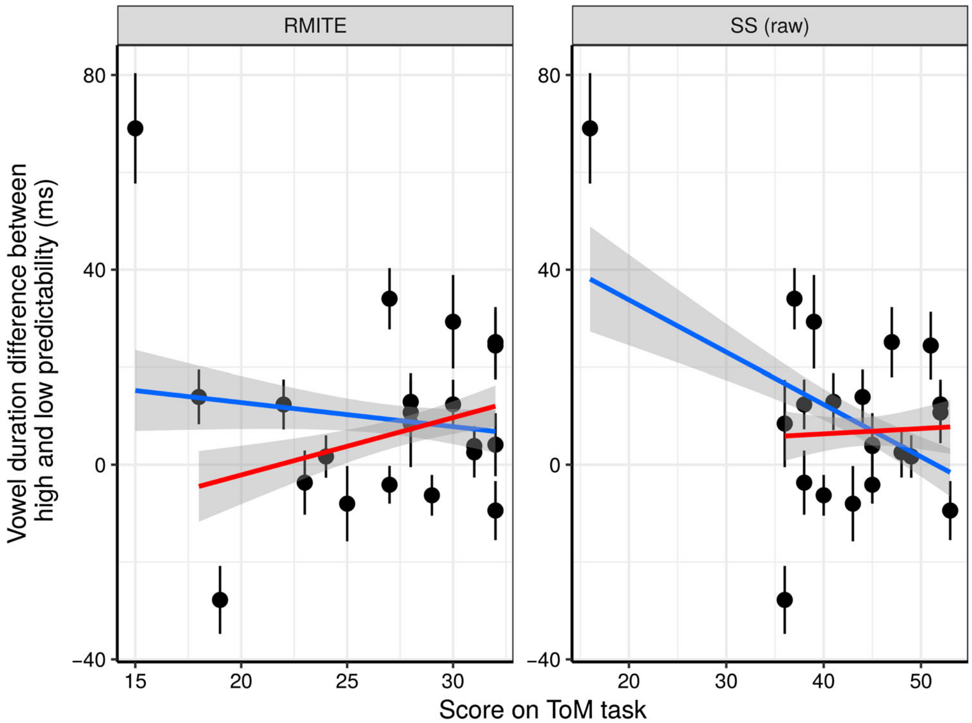

Figure 3. Reduction in vowel duration from low to high predictability conditions as a function of talker RMITE score and raw Strange Stories score. Shorter (red) trendline shows linear fit after removing outlier participant (the point at the top left of the graph). (Colour available online.)

ToM as measured by RMITE score. These effects are not particularly consistent – why should vowel duration be affected for semantic predictability-based reduction but vowel dispersion for second mention reduction? And why the RMITE score and not any of the other measures?

Further, three significant interactions were also observed in the *opposite* direction to this hypothesis: reduction in word duration for predictable words was larger for participants with better ToM as measured by AQ and RMITE scores; and reduction is dispersion for predictable words was larger for participants with better ToM as measured by AQ score. These interactions suggest that talkers with poorer ToM in fact produced larger acoustic differences between the predictability conditions than talkers with better ToM, implying that ToM abilities somehow hinder the effective use of reduction in semantically predictable contexts. This

Table 5. Model outputs for vowel dispersion models in the sentence production task of Experiment 1.

	β	SE	t	p
Intercept	−0.040	0.319	−0.124	0.901
Pred.: Low	0.092	0.038	2.411	0.016
AQ score	−0.007	0.011	−0.696	0.487
Pred:AQ	0.011	0.004	2.946	0.003
	β	SE	t	p
Intercept	−0.036	0.316	−0.113	0.910
Pred.: Low	0.092	0.043	2.144	0.032
RMITE score	0.038	0.021	1.819	0.069
Pred:RMITE	−0.011	0.009	−1.263	0.207
	β	SE	t	p
Intercept	−0.038	0.317	−0.119	0.905
Pred.: Low	0.092	0.041	2.211	0.027
SS (raw)	0.017	0.013	1.271	0.204
Pred:SS(r)	−0.009	0.005	−1.677	0.094
	β	SE	t	p
Intercept	−0.038	0.318	−0.120	0.904
Pred.: Low	0.092	0.044	2.110	0.035
SS (normalised)	0.038	0.050	0.762	0.446
Pred:SS(n)	0.015	0.020	0.768	0.443

Table 6. Model outputs of the word duration models in the paragraph-reading task of Experiment 1.

	β	SE	t	p
Intercept	3.537	13.417	0.264	0.792
Word is pre-pausal	55.320	5.382	10.279	0.000
2nd Mention	−24.487	5.210	−4.700	0.000
AQ score	0.332	0.675	0.492	0.623
Mention:AQ	−0.251	0.286	−0.880	0.379
	β	SE	t	p
Intercept	3.205	13.333	0.240	0.810
Word is pre-pausal	55.433	5.377	10.309	0.000
2nd Mention	−24.441	5.224	−4.679	0.000
RMITE score	−1.454	1.399	−1.039	0.299
Mention:RMITE	−0.179	0.629	−0.285	0.776
	β	SE	t	p
Intercept	3.360	13.414	0.250	0.802
Word is pre-pausal	55.487	5.379	10.315	0.000
2nd Mention	−24.465	5.231	−4.677	0.000
SS (raw)	−0.423	0.872	−0.485	0.628
Mention:SS(r)	0.013	0.399	0.032	0.974
	β	SE	t	p
Intercept	3.397	13.243	0.257	0.798
Word is pre-pausal	55.471	5.376	10.319	0.000
2nd Mention	−24.515	5.250	−4.669	0.000
SS (normalised)	−4.509	3.038	−1.484	0.138
Mention:SS(n)	−0.931	1.371	−0.679	0.497

Table 7. Model outputs of the vowel duration models in the paragraph-reading task of Experiment 1.

	β	SE	t	p
Intercept	3.964	5.443	0.728	0.467
Word is pre-pausal	13.310	2.520	5.283	0.000
2nd Mention	−11.725	2.662	−4.404	0.000
AQ score	0.101	0.324	0.313	0.755
Mention:AQ	−0.014	0.140	−0.097	0.923
	β	SE	t	p
Intercept	3.773	5.358	0.704	0.481
Word is pre-pausal	13.267	2.517	5.272	0.000
2nd Mention	−11.682	2.665	−4.383	0.000
RMITE score	−0.888	0.658	−1.349	0.177
Mention:RMITE	0.028	0.305	0.091	0.928
	β	SE	t	p
Intercept	3.747	5.383	0.696	0.486
Word is pre-pausal	13.285	2.517	5.278	0.000
2nd Mention	−11.677	2.665	−4.381	0.000
SS (raw)	−0.457	0.406	−1.125	0.261
Mention:SS(r)	−0.007	0.192	−0.038	0.970
	β	SE	t	p
Intercept	3.926	5.400	0.727	0.467
Word is pre-pausal	13.322	2.514	5.299	0.000
2nd Mention	−11.739	2.699	−4.349	0.000
SS (normalised)	−1.489	1.500	−0.993	0.321
Mention:SS(n)	−0.698	0.686	−1.017	0.309

state of affairs is not predicted by the listener-oriented model, nor is it predicted by alternative models.

The other main result of the experiment was that, aside from the interaction between dispersion and RMITE, second mention reduction was not observed to interact with ToM skills in any way. All participants, regardless of ToM assessment scores, produced second mention reduction to roughly the same degree.

The unexpected effect of AQ on semantic predictability-based reduction defies both listener-based and non-listener-based understandings of phonetic reduction.

Table 8. Model outputs of the vowel dispersion models in the paragraph-reading task of Experiment 1.

	β	SE	t	p
Intercept	0.051	0.200	0.256	0.798
Word is pre-pausal	0.049	0.047	1.055	0.292
2nd Mention	−0.115	0.026	−4.392	0.000
AQ score	0.001	0.006	0.092	0.927
Mention:AQ	0.001	0.002	0.404	0.687
	β	SE	t	p
Intercept	0.053	0.200	0.267	0.789
Word is pre-pausal	0.050	0.047	1.079	0.281
2nd Mention	−0.115	0.025	−4.667	0.000
RMITE score	0.014	0.013	1.021	0.308
Mention:RMITE	−0.013	0.005	−2.714	0.007
	β	SE	t	p
Intercept	0.055	0.200	0.274	0.784
Word is pre-pausal	0.047	0.047	1.004	0.315
2nd Mention	−0.115	0.025	−4.662	0.000
SS (raw)	0.009	0.008	1.056	0.291
Mention:SS(r)	−0.007	0.003	−2.152	0.032
	β	SE	t	p
Intercept	0.051	0.200	0.255	0.799
Word is pre-pausal	0.050	0.047	1.062	0.288
2nd Mention	−0.115	0.026	−4.464	0.000
SS (normalised)	−0.010	0.030	−0.333	0.740
Mention:SS(n)	−0.010	0.011	−0.947	0.344

One potential recourse is to deny the validity of AQ, and attribute variation in AQ to some other underlying variable. General intelligence, measured via IQ (intelligence quotient) has been shown to weakly correlate with AQ in individuals with autism spectrum disorder (Bishop & Seltzer, 2012) (but cf. Rajkumar, Yovan, Raveendran, & Russel, 2008). Moreover, Hoekstra, Happé, Baron-Cohen, & Ronald's (2010) longitudinal twin study suggested that a small genetic link between autism and intelligence exists, although is little research into such a link in neurotypical populations. It is therefore not inconceivable that the high-AQ participants in this experiment also happened to have high IQ scores.

It is also possible that RMITE is likewise correlated with IQ score. A meta-analysis by Baker, Peterson, Pulos, and Kirkland (2014) of reports of RMITE and IQ scores revealed that general intelligence correlates positively with RMITE score ($r = .24 \pm .06$). This relationship may be due to the reliance on verbal labels in the task. Peterson and Miller (2012) found that individual differences in facial recognition ability could not predict RMITE scores, which underscores the role of language in completing the task (see also Johnston, Miles, & McKinlay, 2008, for a critical review of the linking hypotheses underlying the RMITE task).

These ToM scores being correlated with IQ does not make it any easier to explain or account for the results, as there is likewise no known mechanism linking general intelligence with sensitivity to linguistic predictability. However, if the scores are correlated with IQ, it means that the results cannot be solely attributed to ToM. Measuring general intelligence along with the ToM scores therefore serves as a safeguard to ensure the validity of the conclusions drawn from the results. Experiment 2 therefore replicates the sentence-reading task and includes a task designed to estimate IQ.

2.4.2. Usefulness of individual difference scores

Both the RMITE and AQ scores were observed to be useful in explaining variance in the sentence reading task but the Strange Stories task was not. Part of this discrepancy may be related to task demands. Both the RMITE and AQ tasks require the participant to choose one of four possible responses to a limited task. The Strange Stories task, on the other hand, requires greater levels of heightened attention: participants must read a story, multiple sentences in length, and attend to several details. Their response to the question is entirely free-form: they can respond with a single word or a whole essay, making the range of possible responses effectively infinite. Given these constraints, it is possible that some participants received low scores simply due to putting a low level of effort into their answers, or because they were unsure of the necessary level of detail required in the responses. Scoring the answers is also a somewhat subjective affair, despite the clear rubric, and is a potential source of variation in these scores. Finally, the task was designed for use primarily with children, while the AQ and RMITE task were both designed for adults. The often trivially easy questions in some filler items of the Strange Stories task may have perturbed the adult undergraduate participants. Due to these concerns, the significant amount of experiment time required to administer the task, and the significant amount of researcher time required to score the responses, the Strange Stories task was not used in Experiments 2 or 3.

2.4.3. Consciousness and control

A related question that arises from these results is the extent to which any of these observed phonetic reductions are *controlled* by the talker, or if they are automatic, arising from unconscious processes. There is debate in the speech production literature about the extent to which particular acoustic cues can be controlled versus what can be ascribed to "automatic" effects resulting from the physiology of the speech apparatus (see Solé, 2007, for review). A related area of study is that of "clear speech": speech produced under adverse conditions such as heavy background noise or directed to a hearing-impaired listener (see Smiljanić & Bradlow, 2009, for review). The relationship between such explicit speech control and processes assumed to be "automatic", such as phonetic reduction, is relatively understudied (see also Dijksterhuis & Aarts, 2010; Moors, 2006, for recent reviews of the concept of automaticity in social psychology). There is evidence that performance in a trained motor skill can be negatively affected by focusing attention on components of the action (see Beilock & Carr, 2001, and the references therein); it is thought that the conscious attention overrides the more efficiently organised automatic processes involved. From this perspective, it might be expected that many "automatic" processes fail to occur in clear speech, where the talker is effortfully trying to make their speech more clear. However, in one of the few experimental investigations of these questions, Baker and Bradlow (2009) observed both frequency effects – phonetic reduction on more frequent words relative to less frequent words – and second mention reduction in clear speech, suggesting that the mechanisms of clear speech do not necessarily "override" these processes (see also Burdin, Turnbull, & Clopper, 2015; Clopper, Turnbull, & Burdin, 2018). The relationship between clear speech and semantic predictability-based reduction, however, remains understudied.

Additionally, the literature on clear speech has established that there is considerable variation between talkers in their strategies and effectiveness in implementing clear speech (Ferguson, 2004, 2012). The extent to which these strategies can be related to measurable patterns of individual differences in cognition remains understudied. One hypothesis relating clear speech to theory of mind is that individuals with better ToM produce greater or more effective enhancements in clear speech, due to their ability to put themselves in the shoes of their interlocutor. Individuals with poorer ToM, on the other hand, should implement clear speech in a less effective manner. Experiment 2 considers this possibility by examining interactions between semantic predictability-based reduction, clear speech, and theory of mind.

3. Experiment 2: semantic predictability-based reduction and clear speech

3.1. Introduction

The goals of this experiment were directly motivated by the preceding discussion. It sought to: replicate the unexpected interaction between AQ and word predictability; control for possible variance in IQ, and relate variation in intelligence to variation in the AQ and RMITE scores; and investigate interactions between semantic predictability-based reduction, theory of mind, and speech style. To this end, the experiment involved the AQ and RMITE tasks, an IQ estimation task, and an extended version of the sentence reading task from Experiment 1.

3.2. Methods

3.2.1. Participants

Twenty-five eligible participants completed the task for partial course credit. As with Experiment 1, eligibility was restricted to native monolingual speakers of American English with no history of speech, language, or hearing disorders. None of the participants reported any history of autism-spectrum conditions, nor had any participated in Experiment 1.

3.2.2. Procedure

The same task and materials as in Experiment 1 were used to elicit semantic predictability based reduction. In the plain speech condition, sentences were presented on screen for 3500 ms at a time and participants were directed to read "as if talking to a friend". As in Experiment 1, the sentences were blocked by predictability condition and randomised within each block. The order of blocks was counterbalanced between participants. The clear speech condition followed the plain speech condition: participants were told that they would read the sentences again "as if talking to someone with a hearing impairment or who is a non-native speaker of English." This method of eliciting different speaking styles is well-established in the literature (Turnbull & Clopper, 2013). The clear speech condition was otherwise identical to the plain speech condition, except that the stimuli were presented for 4000 ms at a time rather than 3500 ms,[8] and the exact order of the sentences within each predictability block was re-randomised. Following the reading task, participants completed the AQ questionnaire, the RMITE test, and an IQ assessment (KBIT-2; Kaufman & Kaufman, 2004).

3.2.3. Measurements

As in Experiment 1, the word duration, vowel duration, and midpoint F1 and F2 of the target words were measured, and vowel dispersion calculated.

3.2.4. Analysis

Linear mixed effects regression models were constructed to model each of the acoustic variables – word duration, vowel duration, and vowel dispersion. As in Experiment 1, separate models were constructed for each ToM task. Fixed effects were predictability condition (high versus low, reference: high), speech style (plain versus clear, reference: plain), IQ estimate, and ToM score (AQ or RMITE). All two- and three-way interactions between predictability condition, speech style, and IQ were included, as well as two- and three-way interactions between predictability condition, speech style, and ToM score. Therefore, the fixed effect structure was as follows:

$$\text{predictability} \times \text{style} \times (\text{ToM} + \text{IQ}).$$

All continuous variables were centred around the mean before being entered into the model. Random intercepts for talker and word identity were included, with random slopes for predictability condition and style for both talker and word. As in Experiment 1, influence diagnostics were calculated for each word and each participant in each model. The calculation of *p*-values was the same as that of Experiment 1.

From a total of 4100 tokens (41 words × 2 predictability conditions × 2 styles × 25 talkers), 118 disfluent utterances and misreadings were excluded. Additionally, all formant values more than three standard deviations away from their talker by-vowel means were excluded, as were all duration values more than three standard deviations away from their talker means. This process resulted in a total of 3980 word duration measures, 3970 vowel duration measures, and 3170 vowel dispersion measures being entered into the analysis.

3.3. Results

3.3.1. Individual differences scores

In this sample, AQ scores ranged from 82 to 127, with a mean of 107.80 ($SD = 12.60$); RMITE scores ranged from 19 to 36, with a mean of 27.84 ($SD = 4.31$); and estimated IQ scores ranged from 81 to 133, with a mean of 108.20 ($SD = 13.97$). These AQ and RMITE scores are comparable to those of Experiment 1 in terms of range, central tendency, and dispersion. Table 9 shows a correlation matrix, with *p*-values, between these three variables. The IQ and RMITE scores are positively correlated, consistent with the findings of Baker et al. (2014).

3.3.2. Word duration

Table 10 shows the output for the word duration models from Experiment 2. Two significant simple effects were observed, those of word predictability and speech style. As expected, words in the low-predictability condition were longer ($M = 443$ ms) than those in the high-predictability condition ($M = 444$ ms), confirming that predictability-based reduction occurred. Additionally, words in the clear speech condition were significantly longer ($M = 483$ ms) than those in the plain speech condition ($M = 393$ ms), consistent with decades of previous research on clear speech effects (Smiljanić & Bradlow, 2009).

Talker AQ score was observed to significantly interact with style, such that talkers with a higher AQ score produced shorter word durations in the clear speech condition relative to talkers with a lower AQ score. In other words, the higher AQ talkers (those with a poorer ToM) had a smaller clear speech effect – their words were less enhanced, relative to the plain condition. This effect is visualised in Figure 4. While no other significant interactions were observed, a three-way interaction between style, predictability, and AQ was trending ($p = .085$). This interaction is visualised in Figure 5, where it can be seen that clear speech features

Table 9. Correlation matrix of the individual difference scores in Experiment 2, also including *p*-values for the significance of each pairwise comparison.

	AQ score		RMITE	
	r	(*p*)	*r*	(*p*)
RMITE score	−.206	(.324)		
IQ score	−.022	(.917)	.431	(.031)

Table 10. Outputs for the word duration models for Experiment 2.

	β	SE	t	p
Intercept	−50.231	15.691	−3.201	0.001
Style: Clear	89.231	9.104	9.801	0.000
Pred: Low	9.118	3.904	2.336	0.020
IQ	−0.123	0.542	−0.227	0.821
AQ score	−0.776	0.601	−1.292	0.197
Style:Pred	2.558	3.489	0.733	0.463
Style:IQ	0.345	0.614	0.562	0.574
Pred:IQ	0.317	0.257	1.232	0.218
Style:AQ	−1.452	0.680	−2.135	0.033
Pred:AQ	−0.091	0.285	−0.321	0.748
Style:Pred:IQ	−0.258	0.255	−1.010	0.312
Style:Pred:AQ	0.487	0.283	1.725	0.085
	β	SE	t	p
Intercept	−50.309	15.812	−3.182	0.001
Style: Clear	89.265	9.518	9.379	0.000
Pred: Low	9.171	3.750	2.445	0.015
IQ	0.080	0.621	0.129	0.897
RMITE score	−1.409	2.015	−0.700	0.484
Style:Pred	2.570	3.490	0.736	0.462
Style:IQ	0.181	0.716	0.253	0.801
Pred:IQ	0.191	0.271	0.705	0.481
Style:RMITE	1.445	2.323	0.622	0.534
Pred:RMITE	0.951	0.882	1.078	0.281
Style:Pred:IQ	−0.384	0.282	−1.361	0.173
Style:Pred:RMITE	0.897	0.921	0.974	0.330

a pattern opposite to that of plain speech with regards to interactions between reduction and theory of mind. Since this interaction is not significant, it is not clear how (or whether) to interpret it but the pattern is nevertheless striking.

3.3.3. Vowel duration

Table 11 shows the output for the vowel duration models in Experiment 2. As in the word duration model, word predictability was significant, such that words in the low-predictability condition had longer vowels (M = 220 ms) than words in the high-predictability condition (M = 211 ms). Likewise, speech style was significant, such that words in clear speech had longer vowels (M = 238 ms) than words in plain speech (M = 193 ms). No other significant effects were observed.

3.3.4. Vowel dispersion

Table 12 shows the output for the vowel dispersion models in Experiment 2. Consistent with prior research on clear speech (e.g. Krause & Braida, 2004; Picheny, Durlach, & Braida, 1986), words in plain speech had less disperse vowels (M = 3.11 ERB) than words in clear speech (M = 3.35 ERB). Word predictability was not observed to be significant, although the numeric trend is in the expected direction. In the RMITE model only, speech style and predictability interacted significantly, such that the direction of the predictability effect was reversed in the clear speech style relative to the plain speech style ($M^{high}_{plain} = 3.09$ ERB, $M^{low}_{plain} = 3.12$ ERB, $M^{high}_{clear} = 3.37$ ERB, $M^{low}_{clear} = 3.33$ ERB), conceptually concordant with Baker & Bradlow's (2009) finding of larger mention effects on high-frequency words in plain

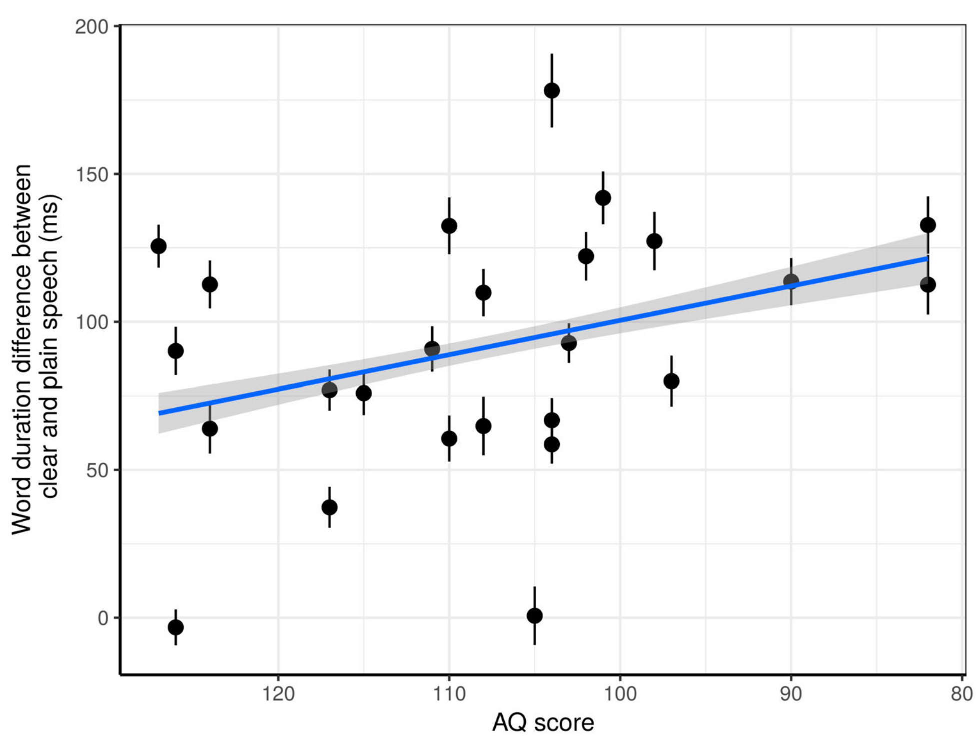

Figure 4. Difference in word duration between clear and plain speech as a function of talker AQ score. Note that the *x*-axis is flipped for consistency with Figure 2, so that participants with poorer ToM cluster to the left of the graph and participants with better ToM cluster to the right of the graph.

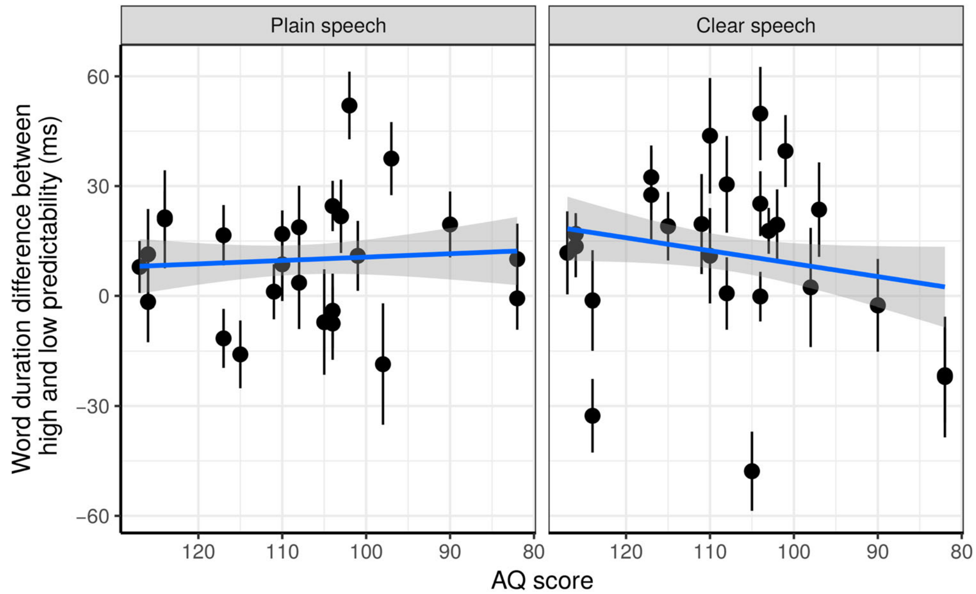

Figure 5. Difference in word duration between predictable and unpredictable words as a function of talker AQ score, split by speech style. Note that the *x*-axis is flipped for consistency with Figures 2 and 4, so that participants with poorer ToM cluster to the left of the graph and participants with better ToM cluster to the right of the graph.

speech relative to clear speech. No other effects or interactions were observed.

3.4. Discussion

The results of this experiment have, in general terms, mirrored those of Experiment 1: no strong evidence for a relationship between ToM scores and reduction was found. Additionally, this experiment controlled for IQ, which failed correlate with any ToM measure, supporting the conclusion from Experiment 1 that the observed phonetic variation is not due to general intelligence. Inconsistencies between the two experiments were observed, however. This experiment did not reveal any unambiguous interactions between ToM scores and reduction, whereas Experiment 1 found significant effects in both directions.

Table 11. Outputs for the vowel duration models for Experiment 2.

	β	SE	*t*	*p*
Intercept	−27.241	8.154	−3.341	0.001
Style: Clear	44.835	5.459	8.213	0.000
Pred: Low	8.737	2.009	4.350	0.000
IQ	−0.216	0.421	−0.513	0.608
AQ score	−0.212	0.467	−0.455	0.649
Style:Pred	0.737	2.332	0.316	0.752
Style:IQ	0.390	0.342	1.141	0.254
Pred:IQ	0.256	0.143	1.788	0.074
Style:AQ	−0.647	0.379	−1.708	0.088
Pred:AQ	0.156	0.158	0.984	0.325
Style:Pred:IQ	−0.172	0.171	−1.011	0.312
Style:Pred:AQ	−0.106	0.189	−0.560	0.576
	β	SE	*t*	*p*
Intercept	−27.239	8.166	−3.335	0.001
Style: Clear	44.829	5.726	7.829	0.000
Pred: Low	8.742	1.958	4.466	0.000
IQ	−0.172	0.468	−0.367	0.714
RMITE score	−0.303	1.518	−0.200	0.842
Style:Pred	0.750	2.332	0.322	0.748
Style:IQ	0.286	0.404	0.708	0.479
Pred:IQ	0.154	0.154	0.999	0.318
Style:RMITE	0.883	1.311	0.674	0.500
Pred:RMITE	0.743	0.503	1.477	0.140
Style:Pred:IQ	−0.145	0.189	−0.769	0.442
Style:Pred:RMITE	−0.189	0.615	−0.308	0.758

Table 12. Outputs for the vowel dispersion models for Experiment 2.

	β	SE	*t*	*p*
Intercept	−0.168	0.205	−0.819	0.413
Style: Clear	0.274	0.048	5.708	0.000
Pred: Low	0.057	0.040	1.408	0.159
IQ	−0.006	0.010	−0.607	0.544
AQ score	−0.002	0.010	−0.164	0.870
Style:Pred	−0.096	0.056	−1.728	0.084
Style:IQ	−0.002	0.004	−0.451	0.652
Pred:IQ	0.002	0.003	0.626	0.531
Style:AQ	−0.006	0.004	−1.567	0.117
Pred:AQ	0.002	0.003	0.600	0.548
Style:Pred:IQ	−0.002	0.004	−0.393	0.694
Style:Pred:AQ	0.004	0.004	0.783	0.434
	β	SE	*t*	*p*
Intercept	−0.164	0.200	−0.820	0.412
Style: Clear	0.292	0.048	6.061	0.000
Pred: Low	0.053	0.039	1.336	0.182
IQ	−0.010	0.010	−0.968	0.333
RMITE score	0.034	0.028	1.203	0.229
Style:Pred	−0.108	0.054	−2.008	0.045
Style:IQ	0.000	0.004	0.030	0.976
Pred:IQ	0.003	0.003	0.947	0.344
Style:RMITE	−0.006	0.011	−0.539	0.590
Pred:RMITE	−0.013	0.010	−1.384	0.166
Style:Pred:IQ	−0.003	0.005	−0.744	0.457
Style:Pred:RMITE	0.009	0.013	0.678	0.498

The clear speech manipulation itself also yielded interesting results. As expected, words in clear speech were longer and had longer and more disperse vowels than words in plain speech. Additionally, talkers with higher AQ scores had smaller temporal clear speech effects – these talkers did not enhance their word durations ($p = .033$) as much as the talkers with lower AQ scores, and the results for vowel duration trended in the same direction ($p = .088$). This finding is suggestive of a role for theory of mind in listener-oriented speech styles: the talkers with poor ToM (high AQ scores) were less skilled at adapting their speech for their (imagined) interlocutor than the talkers with good ToM.[9] Finally, the fact that the predictability by style interactions failed to reach significance for word and vowel duration suggests that predictability effects influence clear speech and plain speech equally. This fact further suggests that these effects are well outside of conscious control. If the explicit control of clear speech were able to "override" the influence of predictability on production – that is, if talkers could effectively engage and disengage the predictability effects at will – then we would expect to observe great variability in clear speech, including the possibility of the absence of predictability effects. The data do not match this prediction, and there is therefore no evidence that these predictability effects are subject to conscious control.

4. General discussion

4.1. Summary of experiments

The experiments reported in this paper investigated the relationship between phonetic reduction and individual variation in theory of mind. Experiment 1 investigated second mention reduction and semantic predictability-based reduction. Second mention reduction was observed in both the temporal and spectral domains: second mentions of words were shorter with less disperse vowels than first mentions. No interactions with ToM skill was observed. Semantic predictability-based reduction was also observed in both the temporal and spectral domains. Inconsistent interactions with ToM skill were observed: for word duration, such that talkers with poorer ToM produced *larger* magnitudes of reduction, while for vowel duration, talkers with poorer ToM produced *smaller* magnitudes of reduction. Taken together these inconsistent effects suggest a limited role for ToM in phonetic reduction.

The results of Experiment 2 provided tentative support for this conclusion: none of the ToM measures were observed to significantly interact with phonetic reduction. Experiment 2 controlled for general intelligence, a possible confound on ToM skill, and also tested for potential interactions in a "clear speech" condition, where participants were talking as if to someone with a hearing impairment or a non-native speaker of English. While all talkers exhibited temporal enhancement in the clear speech condition relative to the plain speech condition, this enhancement was larger for talkers with stronger ToM skills (*viz* lower AQ scores).

4.2. Listener-oriented theories

Turning now to the listener-oriented theories which form the theoretical basis for this study, we address the implications the present results hold for these accounts of phonetic reduction. In the introduction, we outlined the major predictions of a strong interpretation of listener orientation: that ToM should negatively correlate with the extent and consistency of reduction. The data from the current experiments suggest that ToM failed to correlate with effects of semantic predictability-based reduction or second mention reduction. These results are incompatible with the predictions of a listener-oriented account. These results are, however, consistent with "talker-based" or "egocentric" accounts of speech production (Bard et al., 2000; Bard & Aylett, 2005; Keysar & Barr, 2005; Keysar et al., 2000), where phonetic reduction is a function of activation or accessibility within the speech production system: words which are accessible in the discourse context are highly activated, and therefore are produced in a reduced way, regardless of the listener's needs.

Weaker interpretations of listener-oriented accounts exist which do not necessarily require interlocutor-specific knowledge but instead rely on "generic listener" models. Turk (2010, p. 230) noted:

> Crucially, [this] proposal does not require that the speaker necessarily take the listeners into account during the online speaking process. The speaker's language redundancy computation can be made on the basis of his or her own language experience. While not necessarily optimal for the listener, this type of language redundancy computation may represent a reasonable approximation to the language redundancy of the listener. Information about the listeners' knowledge or experience can be incorporated in the computation, but doesn't have to be.

However, this formulation still requires the talker to have some mental model of a knowledge state which is different from their own knowledge state at the point of motor planning. That is, from the perspective of the talker, an "unpredictable" word like *chunks* in the sentence *I did not know about the chunks* is actually quite predictable, because it has already been planned and is central to what they want to say. In order to know that *chunks* is unpredictable, the talker must adopt the perspective of a listener who does not know what the talker is going to say. In other words, the talker must have a theory of mind, despite not explicitly requiring a mental model of the interlocutor in question.[10]

Experiment 1 revealed inconsistent interactions between ToM and phonetic reduction. Assuming for the moment that these interactions represent genuine patterns and not noise, these patterns cannot be accounted for by either a weak listener-oriented theory nor an egocentric theory. It is reasonable to assume that a portion of this difference is due to idiosyncratic individual variability; after all, it is well-established that some people have generally more intelligible and clearer speech than other people, regardless of context (Ferguson, 2004; Picheny, Durlach, & Braida, 1985). While some of this variance in speech production could be due to non-pathological physiological factors (e.g. mild dental malocclusions; Kummer, 2008), it is reasonable to assume that some portion of this variance is due to cognitive and personality factors, such as theory of mind. There is currently no role for these factors in either talker-oriented or listener-oriented theories.

The linking hypothesis of the current study assumes that participants with a weaker ToM are simply slower at modelling the mental states of others. This slowness allows less time for the listener-modelling mechanism to exert influence on phonetic planning, leading to less phonetic adjustment (reduction or enhancement). However, speed/accuracy trade-offs are common in cognitive processes and ToM may be no different. This perspective would predict that participants with weaker ToM have more *variable* phonetic reduction – they may reduce or enhance in inappropriate contexts, due to their incorrect modelling of the listener's mental state (see also Keysar & Barr, 2005). Both predictions are simultaneously possible: that is, weaker ToM leads to less phonetic reduction and more variable phonetic reduction. An additional consideration is whether listener-modelling is only "active" for certain tasks – such as sentence production but not for isolated word production, and so on. Related questions have been raised by Buz and Jaeger (2016) regarding links between articulation and neighbourhood density: it is not necessarily the case that neighbourhood density is a good measure of communicative difficulty, especially in tasks with no sentence context (see also Gahl & Strand, 2016).

Taken together, these considerations suggest that a wholesale rejection of a strong interpretation of the listener-oriented account may be premature. Nevertheless, the results of the current experiments do not lend support to this account's predictions, and as a consequence a non-listener-oriented account (e.g. Bell et al., 2009; Pierrehumbert, 2002) seems preferable.

4.3. What listener?

The experiments reported in this paper each involved an imagined interlocutor rather than a real interlocutor. Talkers were asked to imagine that they were talking to a friend (or, in the clear speech condition of Experiment 2, someone with a hearing impairment) but no actual listener was physically present. Do we expect listener modelling in the absence of a listener?

The literature suggests that both explicit speech style adjustments (Ferguson & Kewley-Port, 2007; Picheny et al., 1986; Smiljanić & Bradlow, 2005) and predictability effects (Baker & Bradlow, 2009; Clopper & Pierrehumbert, 2008; Clopper et al., 2018; Munson & Solomon, 2004) still obtain in these environments. As pointed out by Clopper and Turnbull (2018), participants in laboratory studies are aware that their speech is being recorded and that someone will eventually listen to their speech (see also Wagner, Trouvain, & Zimmerer, 2015). Speech recorded in laboratory settings therefore cannot be regarded to be self-directed or non-communicative. Indeed, the results of laboratory experiments with and without real interlocutors are often indistinguishable – Baese-Berk & Goldrick's (2009) study, featuring a real interlocutor, has been replicated without an interlocutor by Bullock-Rest et al. (2013) and Fox, Reilly, and Blumstein (2015) (among others) with the same pattern of results. This trend is in spite of the fact that other phonetic dimensions, such as coarticulation, can and do differ with respect to the presence or absence of an interlocutor (see e.g. Scarborough & Zellou, 2013). Given these considerations, there is no compelling reason to believe that the results of the present experiment were biased due to the lack of a live interlocutor.

4.4. Measuring theory of mind and other dimensions of individual differences

Experiment 1 used three different tasks to assess theory of mind: the Autism-spectrum Quotient (AQ), the Reading the Mind in the Eyes (RMITE) test, and the Strange Stories (SS) task. In general these measures were weakly correlated with each other: Table 13 depicts correlations among AQ, RMITE, SS, and IQ from Experiments 1 and 2. As the SS task was only used in Experiment 1 and IQ was estimated only in Experiment 2, the values for those correlations are the same as those presented in Tables 2 and 9. It is worth noting that this study was not designed to investigate correlations among these variables, and as such any inferential tests are underpowered. With $N = 25$, as in Experiment 2, assuming a modest correlation of $r = .25$ the power to observe a significant effect is only .218. Pooling the data from Experiments 1 and 2 gives $N = 46$ and again assuming $r = .25$ yields a power of .394. The correlation tables in Tables 2, 9, and 13 should therefore be regarded as suggestive and exploratory.

It was assumed that each of these tasks measures some aspect of ToM, although the present design offers no way to validate this assumption. The AQ assesses a broad spectrum of autistic traits, especially social skills and communication (Stewart et al., 2015). The RMITE is a measure of visual emotional cognition, although it has been argued that it is primarily a linguistic task with limited relationship to emotional cognition (Johnston et al., 2008). Both of these measures showed some degree of interaction with phonetic reduction in the present study.

The Strange Stories task focuses on interpreting and explaining the motivations of characters in stories. It did not significantly correlate with phonetic reduction; it is possible that this failure was due to the simplicity of the questions and stories, which were originally designed for children. Some adult participants in the current experiments may have felt the answers were self-evident and elected not to provide full explanations. The recently developed Strange Stories Film Task (Murray et al., 2017) may be a more appropriate instrument.

The AQ is claimed to assess five specific subscales, based on classic diagnostic criteria of autism: communication, social skills, attention switching, imagination, and attention to detail. It is therefore possible to carry out analyses on particular subscales, rather than the entire aggregate AQ score. However, subsequent psychometric research has failed to reach a consensus on the validity of these subscales, with different factor structures holding for different populations (Austin, 2005; Broadbent, Galic, & Stokes, 2013; do Egito, Ferreira, Gonçalves, & Osório, 2018; Eriksson, 2013; Hoekstra, Bartels, Cath, & Boomsma, 2008; Hoekstra et al., 2011; Hurst, Mitchell, Kimbrel, Kwapil, & Nelson-Gray, 2007; James, Dubey, Smith, Ropar, & Tunney, 2016; Kloosterman, Keefer, Kelley, Summerfeldt, & Parker, 2011; Lau, Kelly, & Peterson, 2013; Lundqvist & Lindner, 2017; Palmer, Paton, Enticott, & Hohwy, 2015; Ruzich et al., 2015; Stewart & Austin, 2009, 2010; Zhang et al., 2016). Examining the subscales for correlations would therefore seem to be an exercise in data fishing rather than confirmatory hypothesis testing.[11]

It is possible that some other correlated dimension of cognitive or personality variation is responsible for the phonetic variation observed in the present data. More generally, the role of this kind of individual variability in speech production and perception remains understudied. Measurable personality traits that have at least some theoretical grounding in being potentially relevant for speech production and perception include interpersonal orientation (Swap & Rubin, 1983) and self-monitoring (Snyder, 1974), which are measures of how one's behaviour relates to others and oneself, respectively. Street and Murphy (1987) offered preliminary evidence that interpersonal orientation influences conversation behaviour, such as speech rate, interruption frequency, and turn duration, suggesting that further investigation of these variables of personal variation is warranted. The self-monitoring scale in particular would appear to have an intuitive link to the phenomenon of sociolinguistic accommodation, although we know of no studies to date which explore this question. A similar question, however, was examined by Aguilar et al. (2016), who investigated the role of rejection sensitivity (Downey & Feldman, 1996) – a measure of the extent to which an individual anticipates, perceives, and negatively reacts to social rejection – in mediating speech accommodation in dyadic interactions. Their results suggested that individuals who are highly sensitive to rejection tend to accommodate more to their conversational partner than less sensitive individuals, although the degree to which the interlocutors' rejection sensitivity were similar or dissimilar also influenced the degree of accommodation. Taken together, Aguilar et al. (2016) and Street and Murphy (1987) offer compelling evidence that personality factors can have a powerful influence on behaviours often considered purely (socio)linguistic. Further questions on this topic for future research include the extent to which these speech behaviours

Table 13. Correlation matrix of the individual difference scores from all three experiments, with p-values and N for each pairwise comparison.

	AQ score			RMITE			IQ score			Raw SS score		
	r	*p*	*N*	*r*	*p*	*N*	*r*	*p*	*N*	*r*	*p*	*N*
RMITE score	−.193	.199	46									
IQ score	−.022	.885	25	.431	.003	25						
Raw SS score	−.383	.009	21	.673	.001	21	–	–	0			
Normalised SS score	.129	.394	21	.163	.279	21	–	–	0	−.240	.108	21

are static functions of personality, and whether they can be influenced by interventions.

5. Conclusion

This paper described three experiments which examined relationships between phonetic reduction and individual variation in theory of mind. Observed relationships were small and in inconsistent directions for different acoustic variables, different measures of theory of mind, and different sources of phonetic reduction. Taken together, these results suggest that a strong interpretation of the listener-oriented account of phonetic reduction enjoys only limited explanatory adequacy.

Notes

1. See also Silverman (2012) on the development of anti-homophony constraints in phonology: "Anti-homophony is thus not an *active* pressure for which there is an abundance of overt evidence. Rather, it is a *passive* result of the pressures that inherently act upon the interlocutionary process." (p. 147, emphasis in original.)
2. Note that weaker interpretations of the listener-oriented account exist which do not necessarily require interlocutor-specific knowledge but instead rely on "generic listener" models (Turk, 2010). It is difficult to reconcile this approach with the classic literature on audience design in collaborative tasks (e.g. Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987), which illustrate that talkers consistently employ listener-modelling in their choice of referring expression. The weaker forms of the listener-oriented accounts do not make explicit the conditions under which we expect talkers to use listener-specific knowledge, making it difficult to test the validity of this account.
3. Although the concept of ToM is usually regarded as originating with Premack and Woodruff (1978), the related concepts of metarepresentation and mentalising have been part of biological thinking for many decades; see chapter 12 of Hobhouse (1901) for an early treatment of the subject.
4. Clopper et al.'s (2017) materials were, in turn, adapted from those of Baker and Bradlow (2009).
5. The original coding system outlined by Baron–Cohen, Wheelwright, Skinner et al. (2001) uses a binary scoring system, collapsing the distinction between "strongly (dis)agree" and "(dis)agree". Following more recent work (e.g. Stevenson & Hart, 2017; Yu, 2010), this study used a Likert-style coding system which takes into account all four values. Possible scores on the AQ measured this way range from 50 to 200.
6. Equivalent rectangular bandwidth (ERB), is a psychoacoustically motivated scale similar to the Bark scale.
7. This correlation is significant even after applying a Bonferroni correction: $\alpha = .05/6 \approx .008 > .001$.
8. Pilot testing revealed that participants found presentations of 3500 ms in the clear speech condition to be "too fast". This lengthened time also implicitly encouraged the participants to speak clearly.
9. Whether or not these adaptations are actually helpful for speech intelligibility is a separate question (see e.g. Ferguson, 2004; Picheny et al., 1985).
10. Indeed, it is likewise difficult to reconcile this weaker interpretation with the existing literature on audience design in collaborative tasks. In a classic task in the action-language tradition, two interlocutors worked together to arrange a set of tangram figures (Clark & Wilkes-Gibbs, 1986). During the course of the task, the interlocutors developed "conceptual pacts" whereby the name of a particular figure was conventionalised, such as "the ice-skater" (Brennan & Clark, 1996). A similar study by Isaacs and Clark (1987) used New York City landmarks instead of tangrams, and had dyads of two experts (native New Yorkers) or an expert and a novice (people who had never visited New York). They found that interlocutors are extremely adept at quickly determining the knowledge level of their dialogue partner, and tailoring their productions accordingly. These findings illustrate that talkers consistently employ listener-modelling in their choice of referring expression. The weak form of the listener-oriented account quoted above holds that talkers can do listener-modelling when needed but that it is not required. This version of the theory does not make explicit the conditions under which we expect talkers to use listener-specific knowledge, which means that testing the validity of this account is difficult.
11. Regardless of the fine structure of the subscales, however, the full AQ score has been demonstrated to be reliable (e.g. Austin, 2005), to compare favourably with other measures of autistic traits (Armstrong & Iarocci, 2013), and to not be easily explainable in terms of personality traits (Wakabayashi, Baron-Cohen, & Wheelwright, 2006), suggesting that the AQ does indeed measure cognitive style rather than simply idiosyncratic personal preferences.

Acknowledgments

I am grateful to Mary Beckman, Cynthia Clopper, Susanne Gahl, Becca Morley, Shari Speer, and two anonymous reviewers for comments and feedback on earlier versions of this paper. This work has benefited from feedback from audiences at Indiana University; the Laboratoire de Sciences Cognitives et Psycholinguistique; the New Zealand Institute for Language, Brain, and Behaviour; the Ohio State Center for Cognitive and Brain Sciences; the University of Edinburgh; the 18th and 19th Mid-Continental Phonetics and Phonology Conferences; the 22nd Manchester Phonology Meeting; and the 88th Annual Meeting of the LSA. I would also like to thank McKenna Reeher and Amy Roberts for their assistance with data collection and coding. All errors are my own.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This research was partially supported by a Presidential Fellowship from the Ohio State University Graduate School.

References

Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., & Bolger, N. (2016). A dyadic perspective of speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality*, *84*(2), 165–177.

Armstrong, K., & Iarocci, G. (2013). The autism spectrum quotient has convergent validity with the social responsiveness scale in a high-functioning sample. *Journal of Autism and Developmental Disorders*, *43*, 2228–2232.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the autism spectrum quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460.

Aylett, M. (2000). *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech* (PhD thesis). University of Edinburgh, Edinburgh.

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

Aylett, M., & Turk, A. E. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*, 3048–3058.

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

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*, 527–554.

Baker, R., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413.

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "reading the mind in the eyes". *Intelligence*, *44*, 78–92.

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language*, *10*, 265–292.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*, 1–22.

Bard, E. G., & Aylett, M. (2005). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge: MIT Press.

Bard, E. G., Lowe, A., & Altmann, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), *Proceedings of Eurospeech '89 (Vol. 2, pp. 573–576) Paris, France*.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–822.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or asperger syndrome. *Visual Cognition*, *4*(3), 311–331.

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

Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, *130*(4), 701–725.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bishop, S. L., & Seltzer, M. M. (2012). Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*, 2354–2363.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(6), 1482–1493.

Broadbent, J., Galic, I., & Stokes, M. A (2013). Validation of autism spectrum quotient adult version in an Australian sample. *Autism Research and Treatment*, *2013*, 984205.

Bullock-Rest, N., Cerny, A., Sweeney, C., Palumbo, C., Kurowski, K., & Blumstein, S. E. (2013). Neural systems underlying the influence of sound shape properties of the lexicon on spoken word production: Do fMRI findings predict effects of lesions in aphasia? *Brain and Language*, *126*, 159–168.

Burdin, R. S., & Clopper, C. G. (2015). Phonetic reduction, vowel duration, and prosodic structure. In Proceedings of the eighteenth international congress of the phonetic sciences. Glasgow, Scotland.

Burdin, R. S., Turnbull, R., & Clopper, C. G (2015). Interactions among lexical and discourse characteristics in vowel production. *Proceedings of Meetings on Acoustics*, *22*, 060005.

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, *31*(3), 404–424.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–30.

Clopper, C. G., Mitsch, J. F., & Tamati, T. N. (2017). Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics*, *60*, 38–59.

Clopper, C. G., & Pierrehumbert, J. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America*, *124*, 1682–1688.

Clopper, C. G., & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, &

M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: Mouton de Gruyter.

Clopper, C. G., Turnbull, R., & Burdin, R. S. (2018). Assessing predictability effects in connected read speech. *Linguistics Vanguard*, *4*(s2), 2017044.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory phonology*, *6*(2), 243–278.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York, NY: Chapman and Hall.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100.

Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology*, *61*, 467–490.

do Egito, J. H. T., Ferreira, G. M. R., Gonçalves, M. I., & Osório, A. A. C. (2018). Factor analysis of the Brazilian version of the adult autism spectrum quotient. *Journal of Autism and Developmental Disorders*, *48*(5), 1847–1853.

Donnarumma, F., Dindo, H., Iodice, P., & Pezzulo, G. (2017). You cannot speak and listen at the same time: A probabilistic model of turn-taking. *Biological Cybernetics*, *111*(2), 165–183.

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology*, *70*(6), 1327–1343.

Engelhardt, P. E., & Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by context? Implications for models of reference production. *Language, Cognition and Neuroscience*, *29*(8), 975–985.

Eriksson, K. (2013). Autism-spectrum traits predict humor styles in the general population. *Humor: International Journal of Humor Research*, *26*(3), 461–475.

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, *48*, 60–75.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*, 2365–2373.

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss. *Journal of Speech, Language, and Hearing Research*, *55*, 779–790.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*, 1241–1255.

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., & Baron-Cohen, S. (2013). Test-retest reliability of the "reading the mind in the eyes" test: A one-year follow-up study. *Molecular Autism*, *4*(1), 33.

Flemming, E. (2010). Modeling listeners: Comments on Pluymaekers et al. and Scarborough. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10*. Berlin: Mouton de Gruyter.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech*, *31*, 307–319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, *26*, 489–504.

Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language*, *37*, 24–40.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, *83*, 97–117.

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

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th annual meeting of the cognitive science society* (pp. 939–944), Washington, DC, USA.

Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, *80*(4), 748–775.

Gahl, S., & Strand, J. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, *89*, 162–178.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35–51.

Genzel, D., & Charniak, E. (2003). Variation of entropy and parse trees of sentences as a function of the sentence number. In *Proceedings of EMNLP* (pp. 65–72), Sapporo, Japan.

Grice, M., Krüger, M., & Vogeley, K. (2016). Adults with asperger syndrome are less sensitive to intonation than control persons when listening to speech. *Culture and Brain*, *4*(1), 38–50.

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

Hawkins, S., & Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, *22*, 493–511.

Hobhouse, L. T. (1901). *Mind in evolution*. London: Macmillan.

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566.

Hoekstra, R. A., Happé, F. G. E., Baron-Cohen, S., & Ronald, A. (2010). Limited genetic covariance between autistic traits and intelligence: Findings from a longitudinal twin study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153B, 994–1007.

Hoekstra, R. A., Vinkhuzen, A. A. E., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., … van der Sluis, S. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short). *Journal of Autism and Developmental Disorders*, *41*, 589–596.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Hunnicutt, S. (1985). Intelligibility versus redundancy – conditions of dependency. *Language and Speech*, *28*(1), 47–54.

Hunnicutt, S. (1987). Acoustic correlates of redundancy and intelligibility. *Quarterly Progress Status Report, Speech Transmission Lab*, *28*(2–3), 7–14.

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the autism spectrum quotient (AQ) in a non-clinical sample. *Personality and Individual Differences*, *43*, 1938–1949.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26–37.

Jaeger, T. F (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*, 230.

Jaeger, T. F., & Tily, H. (2011). On language "utility": Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 323–335.

James, R. J., Dubey, I., Smith, D., Ropar, D., & Tunney, R. J. (2016). The latent structure of autistic traits: A taxometric, latent class and latent profile analysis of the adult autism spectrum quotient. *Journal of Autism and Developmental Disorders*, *46*(12), 3712–3728.

Johnston, L., Miles, L., & McKinlay, A. (2008). A critical review of the eyes test as a measure of social-cognitive impairment. *Australian Journal of Psychology*, *60*(3), 135–141.

Jolliffe, T., & Baron-Cohen, S. (1999). The strange stories test: A replication with high-functioning adults with autism or Asperger syndrom. *Journal of Autism and Developmental Disorders*, *29*(5), 395–406.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*, 311–325.

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, *30*(1–2), 103–119.

Kaiser, E., Li, D. C.-H., & Holsinger, E. (2011). Exploring the lexical and acoustic consequences of referential predictability. In I. Hendrickx A. Branco, S. L. Devi, & R. Mitkov (Eds.), *Anaphora processing and applications*, Number 7099 in Lecture Notes in Artificial Intelligence (pp. 171–183). Heidelberg: Springer.

Kaland, C., Krahmer, E., & Swerts, M. (2014). White bear effects in language production: Evidence from the prosodic realization of adjectives. *Language and Speech*, *57*(4), 470–486.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*(5), 1337–1351.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test* (2nd ed.). Circle Pines, MN: American Guidance Service.

Keysar, B., & Barr, D. J. (2005). Coordination of action and belief in communication. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.

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

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Kloosterman, P. H., Keefer, K. V., Kelley, E. A., Summerfeldt, L. J., & J. D. A. Parker (2011). Evaluation of the factor structure of the autism-spectrum quotient. *Personality and Individual Differences*, *50*, 310–314.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *Journal of the Acoustical Society of America*, *115*(1), 362–378.

Kummer, A. W. (2008). *Cleft palate & craniofacial anomalies: Effects on speech and resonance*. Clifton Park, NY: Cengage Learning.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*(8), 1179–1146.

Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(3), 829–843.

Lau, W. Y. P., Kelly, A. B., & Peterson, C. C. (2013). Further evidence on the factorial structure of the autism spectrum quotient (AQ) for adults with and without a clinical diagnosis of autism. *Journal of Autism and Developmental Disorders*, *43*, 2807–2815.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Proceedings of the twentieth annual conference on neural information processing systems* (pp. 849–856), Vancouver, BC, Canada.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–187.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*, 551–556.

Lindblom, B. (1990a). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer.

Lindblom, B. (1990b). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication*, *6*(4), 220–230.

Loy, A., & Hofmann, H. (2014). HLMdiag: A suite of diagnostics for hierarchical linear models in R. *Journal of Statistical Software*, *56*(5), 1–28.

Lundqvist, L.-O., & Lindner, H. (2017). Is the autism-spectrum quotient a valid measure of traits associated with the autism spectrum? A Rasch validation in adults with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *47*(7), 2080–2091.

Mielke, J., Nielsen, K., & Magloughlin, L (2013). Phonetic imitation by individuals with autism spectrum disorders: Investigating the role of procedural and declarative memory. *Proceedings of Meetings on Acoustics*, *19*, 060142.

Moon, S.-J. (1995). A review of current theories concerning phonetic invariance: With some implications from clear speech. *Language Research*, *31*(3), 481–500.

Moore-Cantwell, C. (2013). Syntactic predictability influences duration. *Proceedings of Meetings on Acoustics*, *19*, 060206.

Moors, A. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058.

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., … Happé, F. (2017). A new test of advanced theory of mind: The "strange stories film task" captures social processing differences in adults with autism spectrum disorders. *Autism Research*, *10*(6), 1120–1132.

Nishiyama, T., Suzuki, M., Adachi, K., Sumi, S., Okada, K., Kishino, H., … Kanne, S. M. (2014). Comprehensive comparison of self-administered questionnaires for measuring quantitative autistic traits in adults. *Journal of Autism and Developmental Disorders*, *44*, 993–1007.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: Mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, *43*(B), 95–106.

Palmer, C. J., Paton, B., Enticott, P. G., & Hohwy, J. (2015). "Subtypes" in the presentation of autistic traits in the general adult population. *Journal of Autism and Developmental Disorders*, *45*(5), 1291–1301.

Pate, J. K., & Goldwater, S. (2011). Predictability effects in adult-directed and infant-directed speech: Does the listener matter? In *Proceedings of the 33rd annual meeting of the cognitive science society*, Boston, MA, USA.

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1–17.

Peterson, E., & Miller, S. F. (2012). The eyes test as a measure of individual differences: How much of the variance reflects verbal IQ? *Frontiers in Psychology*, *3*, 220.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 93–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434–445.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII*. Berlin: Mouton de Gruyter.

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge: MIT Press.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?*Behavioral and Brain Sciences*, *1*(4), 515–526.

Qian, T., & Jaeger, T. F. (2012). Cue effectiveness in communicatively efficient discourse production. *Cognitive Science*, *36*(7), 1312–1336.

Rajkumar, A. P., Yovan, S., Raveendran, A. L., & Russel, P. S. S (2008). Can only intelligent children do mind reading: The relationship between intelligence and theory of mind in 8 to 11 years old. *Behavioral and Brain Functions*, *4*, 51.

Ramscar, M., & Baayen, R. H (2013). Production, comprehension, and synthesis: A communicative perspective on language. *Frontiers in Psychology*, *4*, 233.

Rosa, E. C., Finch, K. H., Bergeson, M., & Arnold, J. E. (2015). The effects of addressee attention on prosodic prominence. *Language and Cognitive Processes*, *30*(1–2), 48–56.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the autism-spectrum quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6* (1), 2.

Sasisekaran, J., & Munson, B. (2012). Effects of repeated production on vowel distinctiveness within nonwords. *Journal of the Acoustical Society of America*, *131*(4), EL336–EL341.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of the Acoustical Society of America*, *134*(5), 3793–3807.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*, 1–24.

Senju, A., Southgate, V., White, S. J., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome. *Science*, *325*(5942), 883–885.

Shields, L. W., & Balota, D. A. (1991). Repetition and associative context effects in speech production. *Language and Speech*, *34*, 47–55.

Silverman, D. (2012). *Neutralization*. Cambridge: Cambridge University Press.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America*, *118*(3), 1677–1688.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*(4), 526–537.

Solé, M.-J. (2007). Controlled and mechanical properties in speech: A review of the literature. In M.-J. Solé, P. S. Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology*. Oxford: Oxford University Press.

Stevenson, J. L., & Hart, K. R. (2017). Psychometric properties of the autism-spectrum quotient for assessing low and high levels of autistic traits in college students. *Journal of Autism and Developmental Disorders*, *47*(6), 1838–1853.

Stewart, M. E., Allison, C., Baron-Cohen, S., & Watson, R. (2015). Investigating the structure of the autism-spectrum quotient using Mokken scaling. *Psychological Assessment*, *27*(2), 596.

Stewart, M. E., & Austin, E. J. (2009). The structure of the austim-spectrum quotient (AQ): Evidence from a student sample in Scotland. *Personality and Individual Differences*, *47*, 224–228.

Stewart, M. E., & Austin, E. J (2010). Corrigendum to "The structure of the autism-spectrum quotient (AQ): Evidence from a student sample in Scotland" [Personality and Individual Differences 47 (2009) 224–228]. *Personality and Individual Differences*, *48*, 88.

Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with "autistic" traits. *Cognition*, *109*, 157–162.

Street, R. L., & Murphy, T. L. (1987). Interpersonal orientation and speech behavior. *Communication Monographs*, *7*, 42–62.

Swap, W. C., & Rubin, J. Z. (1983). Measurement of interpersonal orientation. *Journal of Personality and Social Psychology*, *44*(1), 208–219.

Tily, H., & Kuperman, V. (2012). Rational phonological lengthening in spoken Dutch. *Journal of the Acoustical Society of America*, *132*(6), 3935–3940.

Turk, A. E. (2010). Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology*, *1*(2), 227–262.

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

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*, 445–472.

Turnbull, R. (2015a). *Assessing the listener-oriented account of predictability-based phonetic reduction* (PhD thesis). Ohio State University, Columbus, OH.

Turnbull, R. (2015b). Patterns of individual differences in reduction: Implications for listener-oriented theories. In *Proceedings of the eighteenth international congress of the phonetic sciences*. Glasgow, Scotland.

Turnbull, R. (2017a). The phonetics and phonology of lexical prosody in San Jerónimo Acazulco Otomi. *Journal of the International Phonetic Association*, *47*(3), 251–282.

Turnbull, R. (2017b). The role of predictability in intonational variation. *Language and Speech*, *60*(1), 123–153.

Turnbull, R. (2018). Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguistics Vanguard*, *4*(s2), 20170033.

Turnbull, R., Burdin, R. S., Clopper, C. G., & Tonhauser, J. (2015). Contextual information and the prosodic realization of focus: A cross-linguistic comparison. *Language, Cognition and Neuroscience*, *30*(9), 1061–1076.

Turnbull, R., & Clopper, C. G. (2013). Effects of semantic predictability and dialect variation on vowel production in clear and plain lab speech. *Proceedings of Meetings on Acoustics*, *19*, 060116.

Turnbull, R., Seyfarth, S., Hume, E., & Jaeger, T. F. (2018). Nasal place assimilation trades off inferrability of both target and trigger words. *Laboratory Phonology*, *9*(1), 15.

Vajrabhaya, P., & Kapatsinski, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. In *Proceedings of the seventeenth international congress of the phonetic sciences*, Hong Kong.

Van Son, R., & Pols, L. C. W. (2003). Information structure and efficiency in speech production. In *Proceedings of interspeech* 2003, Geneva, Switzerland.

Van Son, R., & Van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*, 100–123.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "reading the mind in the eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, *18*(4), 326–354.

Wagner, P., Trouvain, J., & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics*, *48*, 1–12.

Wakabayashi, A., Baron-Cohen, S., & Wheelwright, S. (2006). Are autistic traits an independent personality dimension? A study of the autism-spectrum quotient (AQ) and the NEO-PI-R. *Personality and Individual Differences*, *41*, 873–883.

Ward, N. A. (2013). *The role of modality and register in imitation by adults and children* (PhD thesis). University of California, Los Angeles.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic tac toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*, 1548–1557.

White, S. J., Hill, E. L., Happé, F. G. E., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, *80*(4), 1097–1117.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE*, *5*(8), e11950.

Yu, A. C. L. (2016). Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America*, *139*(4), 1672–1690.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, *8*(9), e74746.

Yu, A. C. L., Grove, J., Martinović, M., & Sonderegger, M. (2011). Effects of working memory capacity and "autistic" traits on phonotactic effects in speech production. In *Proceedings of the seventeenth international congress of the phonetic sciences*, Hong Kong.

Zhang, L., Sun, Y., Chen, F., Wu, D., Tang, J., Han, X., … Wang, K. (2016). Psychometric properties of the autism-spectrum quotient in both clinical and non-clinical samples: Chinese version for mainland China. *BMC Psychiatry*, *16*(1), 213.