Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development

Joseph V. Casillas1, Juan José Garrido Pozú1, Kyle Parrish1, Laura Fernández Arroyo1, Nicole Rodríguez2, Robert Esposito3, Isabelle Chang1, Kimberly Gómez1, Gabriela Constantin-Dureci1, Jiawei Shao1, Iván Andreu1, & Katherine Taveras1

1 Rutgers University

2 Poznan

3 Duo-lingo

Author note

We express our gratitude to all the haters.

The authors made the following contributions. Joseph V. Casillas: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Juan José Garrido Pozú: Conceptualization, Writing - Review & Editing; Kyle Parrish: Conceptualization, Writing - Review & Editing; Laura Fernández Arroyo: Conceptualization, Writing - Review & Editing; Nicole Rodríguez: Conceptualization, Writing - Review & Editing; Robert Esposito: Conceptualization, Writing - Review & Editing; Isabelle Chang: Conceptualization, Writing - Review & Editing; Kimberly Gómez: Conceptualization, Writing - Review & Editing; Gabriela Constantin-Dureci: Writing - Review & Editing; Jiawei Shao: Writing - Review & Editing; Iván Andreu: Writing - Review & Editing; Katherine Taveras: Writing - Review & Editing.

Correspondence concerning this article should be addressed to Joseph V. Casillas, Rutgers University - Department of Spanish and Portuguese, 15 Seminary place, New Brunswick, NJ 08904, USA . E-mail: [joseph.casillas@rutgers.edu](mailto:joseph.casillas@rutgers.edu)

Abstract

The present study investigates the interplay between proficiency and individual pragmatic skills in the process of learning a new language. Notably, we focus on the role of empathy in the development of second language (L2) prosody by analyzing the perception and processing of intonation in questions and statements in L2 Spanish. The construct empathy has been shown to influence native language processing in how listeners interpret intonation and meaning when words are ambiguous (Esteve-Gibert et al., 2020). We extend this research to L2 acquisition in order to determine if individual differences in pragmatic skills affect the development of intonation in L2 processing and sentence comprehension. A total of 225 adult L2 Spanish learners (L1 English) from the Northeastern United States completed a two-alternative forced choice (2AFC) task in which they listened to four utterance types and categorized them as either questions or statements. We used Bayesian multilevel regression and Drift Diffusion modeling to analyze the 2AFC data as a function of proficiency and empathy scores. The results showed that learner response accuracy and sensitivity to intonation improved as a function of proficiency. Importantly, higher empathy scores were positively correlated with higher accuracy and processing efficiency in identifying polar interrogatives. Higher empathic individuals, in comparison with lower empathic individuals, appear to be more sensitive to intonation cues in the process of forming sound-meaning associations. The results motivate the inclusion of measures of pragmatic skills, such as empathy, to better account for intonational meaning processing and sentence comprehensionas in second language acquisition.

*Keywords:* Second language acquisition, Intonation, Empathy, Prosody

*Word count:* X

Using intonation to disambiguate meaning: The role of empathy and proficiency in L2 perceptual development

A fundamental difficulty of speech comprehension is that listeners can come to understand different messages when presented with the same linguistic information (Cain, Oakhill, & Lemmon, 2004). This can be especially problematic when one begins the endeavor of learning a new language. In particular, it is common for second language (L2) learners to struggle with intonation—i.e., the melodic contour of an utterance—in the target language (Trofimovich & Baker, 2006). The difficulties associated with intonation can result in comprehension and communication mishaps because the tune is associated not only with linguistic information, e.g., utterance type or syntactic constituency, but also pragmatic information, e.g., polite discourse (Astruc, Vanrell, & Prieto, 2016), bias, or presupposition (Henriksen, Armstrong, & Garcı́a-Amaya, 2016). The present study investigates how the comprehension of intonation develops in adult L2 learners.

Recent research on monolingual populations suggests that individual differences in pragmatic skills, such as empathy, may play a role in meaning disambiguation (Bishop & Kuo, 2016; Esteve-Gibert, Portes, Schafer, Hemforth, & D’Imperio, 2016; Esteve-Gibert et al., 2020; Orrico & D’Imperio, 2020). Concretely, higher empathy individuals, in comparison with lower empathy individuals, appear to be more sensitive to the intonational cues of speech during the process of forming sound-meaning associations. Furthermore, increased attention has been given to how individual differences in learner backgrounds play a role in the process of L2 acquisition (Hu et al., 2013; Liu, 2017; Rota & Reiterer, 2009). The present study contributes to these lines of research by examining how individual differences in pragmatic skills affect the development and processing of intonation during sentence comprehension. Specifically, we investigated the interplay between language proficiency and an individual pragmatic skill (empathy) when learning an L2. We focus on the role of empathy in the development of L2 prosody by analyzing the perception and processing of intonation in questions and statements in L2 Spanish.

## Background and motivation

**L2 acquisition of prosody**. The difficulties associated with learning an additional language in adulthood are numerous. More often than not the focus falls on individual sounds, or segments, though we know that adults who learn an L2 are faced with suprasegmental challenges as well. Concretely, L2 learners often struggle with intonation, i.e., melodic variation at the utterance level. In normal discourse, speakers use intonation to indicate syntactic structure, whether an utterance is a question or a statement, to focus constituents, as well as to convey affective meaning. Notably, the manner in which intonation is mapped to meaning is often language-specific. For these reasons, the development of L2 intonation represents a facet of L2 phonological learning that often results in comprehension and communication difficulties (Trofimovich & Baker, 2006).

Intonation has a semantic function and through adequate cognitive decoding of of the signal a listener can interpret the intention of a given utterance. For example, an intonational contour can indicate to a listener whether the utterance of an interlocutor is declarative or interrogative in nature. As touched upon above, a speaker can use prosody to signal numerous additional pragmatic functions as well. Given the rich variation in pragmatic uses, it is unsurprising that interpreting and decoding intonational contours in a non-native language represents a particularly challenging aspect of speech comprehension for the language learner. Moreover, the use of L1 prosodic features when speaking the target language can result in misunderstandings due to the fact that the same prosodic features may convey different linguistic and paralinguistic meaning in the target language (Chen, 2005; Cruz-Ferreira, 1987, 1987; Mennen, 2007; Pickering, 2001). As noted by Levis (2016) prosody is also “[…] critical for L2 pronunciation because it plays a major role in cementing social bonds as a key marker of social identity” (p. 154).

For learners interested in obtaining native-like pronunciation, intonation is particularly relevant, as prosodic features have been found to be important cues in the perception of non-target-like accents, above and beyond other features of language (Jilka, 2000; Munro, 1995; Pettorino, De Meo, & Vitale, 2014). Nonetheless, intonation is not traditionally taught in the L2 classroom, perhaps because it not common knowledge that proper control of prosody allows the learner not only to produce speech that is more intelligible, but also to comprehend speech in varied communicative settings (de-la-Mota, 2019). Primary focus is generally placed on syntax and morphology, with target language phonology receiving much less, if any, attention (Rao, 2019). When target language pronunciation is addressed, it often focuses on segmental elements (de-la-Mota, 2019), despite the fact that merely being intelligible at the segmental level does not necessarily imply one will be pragmatically understood. As a result, some research has found that intonation is one of the last aspects of L2 phonology that learners acquire (e.g., Kvavik & Olsen, 1974).

Research on L2 intonation has been concerned primarily with speech production. Learner difficulties tend to be ascribed to L1 transfer, and models of L2 phonology by and large focus on the speech segment, as in the Speech Learning Model revised (Flege & Bohn, 2021), or contrasts between segments, i.e., PAM-L2, L2LP (Best & Tyler, 2007; Van Leussen & Escudero, 2015, respectively). Theoretical work that specifically considers prosody in the acquisition of L2 phonology is virtually non-existent, though some researchers have considered how the aforementioned models might account for suprasegmental phenomena (See Trofimovich & Baker, 2006). Nevertheless, a dearth of knowledge remains with regard to how perception of intonation develops in L2 learning, and even less is known about how individual pragmatic differences account for learner outcomes. The purpose of the present project is to address this gap in the literature by examining the perception and processing of intonation during adult L2 phonological acquisition.

**Acquisition of Spanish prosody**. As with all phonetic phenomena, a lack of invariance in the acoustic content of prosodic realizations also increases the difficulty of the learners task. Beyond the level of the individual, however, dialectal differences can account for more difficulties. Spanish is extensively spoken across the world, with relatively small geolectal differences between varieties when compared with other languages, such that speakers from distinct regions can still generally understand each other. That being said, phonetic variation is abundant. For instance, depending on the variety, for the same utterance type a pitch accent may be realized differently with regard to pitch movement and/or syllable duration. Prosodic strategies can be different altogether. Consider absolute interrogatives in Caribbean and Argentine Spanish, which are produced with a nuclear hat pattern in the former and a final falling F0 contour in the latter, both of which differ from the more common final rise found in many other varieties (See Hualde & Prieto, 2015).

Previous research on the acquisition of Spanish prosody has primarily focused on the production of statements and questions, particularly in the study abroad context, using a pre-, post-test designs (Craft, 2015; See Henriksen, Geeslin, & Willis, 2010; Thornberry, 2014; Trimble, 2013a, among others). Though the degree of improvement is variable based on a myriad of factors—such as context formality (Trimble, 2013a), use of Spanish (Henriksen et al., 2010; Trimble, 2013a), social integration (Trimble, 2013a), or the development of meaningful social relationships with native speakers (Thornberry, 2014)—this line of research suggests that learners gradually acquire target-like intonation as they gain experience in the L2. There is a paucity of research on the perception of Spanish intonation, but limited studies corroborate the general finding in speech production that mastery is indeed possible for adult learners (Brandl, González, & Bustin, 2020; Nibert, 2005, 2006; Trimble, 2013b). For instance, Trimble (2013b) examined the perception of intonational cues in statements and absolute interrogatives in L1 English L2 Spanish adult learners that had studied abroad in Venezuela, Spain, or not at all. Using a gating task, Trimble (2013b) found that intonational cues that were absent from participants’ L1 were difficult to perceive, though learners were more accurate with statements than questions, and that familiarity with the target variety improved accuracy. The investigation lends support to the general notion that the L2 intonation system develops in tandem with proficiency in Spanish, which was positively correlated with time spent studying abroad.

In a similar vein, Brandl et al. (2020) also investigated the perceptual development of intonation in questions and statements in L2 Spanish. Specifically, Brandl et al. (2020) examined the effect of L2 proficiency on the perception of broad-focus and narrow-focus declaratives and polar and absolute interrogatives in adult L2 learners of Spanish. The learners completed an AX discrimination task in which they were presented audio and visual stimuli in matched and mismatched conditions. The participants’ task was to determine whether the sentence presented aurally was the same as the sentence presented visually. Brandl et al. (2020) found that perception and processing of L2 intonation improved in conjunction with proficiency in Spanish, though it was conditional on the utterance type, with polar (‘yes/no’) interrogatives being more difficult to process and acquire when compared with simple statements. The authors concluded that perception of L2 intonation develops gradually in conjunction with L2 proficiency.

To summarize, the extant literature suggests that mastery of L2 perception of intonation seems feasible for adult learners, as processing speed and accuracy both improve as L2 proficiency increases. That being said, some utterances present more difficulties than others. Furthermore, familiarity with the L2 variety has a positive impact on learner outcomes, which is particularly relevant given the rich phonetic variability attested in Spanish prosody. Much less is known regarding how perceptual development is modulated by individual differences, such as those related to pragmatic skill.

**Empathy and pragmatic skill**. The construct empathy refers to one’s ability to infer the intentions of others. It is associated with understanding the feelings and emotions of those with whom one interacts (Baron-Cohen & Wheelwright, 2004). Research on empathy has associated the construct with Theory of Mind and perspective-taking (Baron-Cohen, 2011; Carruthers, 2009; Frith & Frith, 2003). Importantly, in recent years empathy has served a proxy for investigating individual pragmatic skill.

Studies on monolingual populations suggest that individual pragmatic skills correlate with variability in semantic/pragmatic interpretation of ambiguous linguistic items, with more pragmatic individuals preferring pragmatically enriched interpretations and less pragmatic individuals preferring literal/semantic interpretations (e.g., Degen & Tanenhaus, 2016; Nieuwland, Ditman, & Kuperberg, 2010). In addition, more pragmatically skilled individuals tend to rely on different phonetic cues to parse syntactically ambiguous sentences with regard to less pragmatically skilled individuals (Bishop & Kuo, 2016). Thus, one possibility is that variability in intonation processing is also linked to individual differences in pragmatic skills. A series of studies has investigated how empathy influences language processing in monolingual populations (Esteve-Gibert et al., 2016, 2020; Orrico & D’Imperio, 2020). This work operationalizes the construct empathy as a pragmatic skill and has focused on it as a source of individual differences.

For instance, Esteve-Gibert et al. (2020) examined how listeners with different levels of empathy interpreted intonation and meaning in contexts in which a temporary lexical ambiguity could only be resolved through intonation. Empathy was measured using a self-report Empathy Quotient (EQ) questionnaire (Baron-Cohen & Wheelwright, 2004) and participants were partitioned into groups corresponding with low or high empathy. Esteve-Gibert et al. (2020) tested French monolinguals in a visual-world paradigm eye-tracking task that resembled a card guessing game. Target objects were homophones in French (e.g., *cane*, Eng. “female duck”; *canne*, Eng. “walking stick”). Esteve-Gibert et al. (2020) found that processing of the lexical ambiguity (the homophones *cane*/*canne*) was modulated by empathy level when intonation was the only cue available. Specifically, highly empathic individuals varied their looking behavior as a function of intonational cues while less empathic individuals did not. That is, higher empathy individuals, in comparison with lower empathy individuals, were found to be more sensitive to intonation cues in the process of forming sound-meaning associations.

In short, individuals with more pragmatic skill (higher empathy) appear to be able to use intonation to resolve temporary lexical ambiguity that can lead to confirmatory vs. contrasting interpretations. This research underscores the importance of considering individual pragmatic differences when examining intonational meaning processing and sentence comprehension. To the best of our knowledge, no studies have explored the construct empathy as it pertains to L2 perceptual development. Thus, we extend this research to the SLA context in order to determine if individual differences in this pragmatic skill affect the development of intonation in L2 processing and sentence comprehension.

## The present study

We investigate how proficiency and empathy are related to the development of L2 prosody by analyzing the perception of intonation in questions and statements in L2 Spanish. This study was pre-registered on the Open Science Framework[[1]](#footnote-22) and designed to address the following research questions:

1. Is perceptional development in L2 Spanish modulated by proficiency and intonation type (i.e., Brandl et al., 2020)?
2. Do pragmatic skills—specifically, empathy—modulate the rate of development in L2 prosody?
3. Does speaker variety affect perception accuracy and processing speed?

Regarding RQ1, we hypothesize that accuracy will increase and processing time will decrease as a function of proficiency and intonation type. As shown in previous studies, total interrogatives (i.e., yes-no questions) will present the most difficulty for L2 learners of Spanish, followed by partial interrogatives (i.e., wh-questions) and declarative broad focus and narrow focus statements. Based on the findings of Esteve-Gibert et al. (2020), we posit that prosodic development will occur sooner and at a faster rate in higher empathy individuals (RQ2). In this operationalization, ‘sooner’ refers to lower proficiency levels in a cross-sectional design, that is, at an earlier developmental stage when compared with lower empathy individuals. Finally, with regard to RQ3, based on tentative findings from native speaker pilot data, we hypothesize that, overall, L2 learners will have the most difficulty (lower accuracy, slower response time) with the Cuban variety.

This project presents a conceptual replication of Brandl et al. (2020) in that we employ a similar experimental paradigm using similar stimuli in order to analyze the relationship between proficiency and L2 perception of intonation. We extend this research by taking into account pragmatic skill, specifically empathy, in L2 sentence processing. Importantly, this research builds on recent studies looking at the role of individual pragmatic skills in language processing and extends them to the field of SLA.

# Method

## Participants

Two hundred twenty-five individuals completed a two-alternative forced choice (2AFC) task in which auditory stimuli were identified as being questions or statements. Participants were recruited using the Prolific.ac online experimental platform and were compensated at a rate of $9.52 per hour for their time. We estimated the task would take approximately 15 minutes to complete, thus each participant was paid $2.70 for completing all three tasks. The mean time to completion was approximately 13 minutes. The pool of participants was filtered using criteria set in Prolific.ac to insure participants self-reported as being L1 English speakers born, raised, and currently living in the Northeastern US with no knowledge of any languages other than English or Spanish. They reported no hearing difficulties and were required to use headphones on a desktop computer. Upon beginning the experiment, all participants responded to the following screening questions: 1) What part of the US are you from? 2) At what age did you begin learning Spanish? 3) Are you proficient in any languages other than English/Spanish? Additionally, participants responded to the prompt “I am most familiar with Spanish from…” and using a pull-down window they selected a variety of Spanish or “I am not familiar with any variety of Spanish”. We excluded data from any participant that responded that they were not from the US Northeast, that they began learning Spanish before the age of 13, or that they were proficient in a language other than English/Spanish. Participants responding categorically across all trials were also excluded. In sum, participants were adult native speakers of American English with varied levels of proficiency in Spanish, ranging from functionally monolingual to highly proficient. All participants with knowledge of Spanish were adult L2 learners, operationally defined as having begun the endeavor of learning Spanish after the age of 13.

## Tasks

The study consisted of three tasks: a 2AFC task, a lexical decision vocabulary assessment, and a lykert-type questionnaire to assess empathy. The tasks were programmed in Python using Pyschopy3 (Peirce et al., 2019) and presented online via Pavlovia. All code and materials used to generate the tasks are freely available on the Open Science Framework (<https://osf.io/dh4zp/?view_only=162d6d13e5814417bcb9de349f18cb62>).

**2AFC**. In the 2AFC task participants were presented with an audio file containing a statement (declarative: broad focus or narrow focus) or a question (yes-no or wh-). Their task was to determine, as quickly and as accurately as possible, if the utterance they heard was a question or a statement. Specifically, they responded to an on-screen prompt asking “Is this a question?” using the keyboard. To respond participants typed ‘1’ for ‘yes’ (i.e., “yes, this is a question”) and ‘0’ for ‘no’ (i.e., “no, this is not a question”).

The auditory stimuli consisted of 64 critical items, 16 of each utterance type. To generate the stimuli, we recorded native Spanish speakers of eight different varieties (Cuban, standard Peninsular, Andalusian, Puerto Rican, Chilean, Argentinean, Mexican, and Peruvian). The eight native speakers all produced the same 64 critical items. All utterances were segmented using Praat (Boersma & Weenink, 2018) and normalized for peak intensity. The 2AFC task included 64 trials in which the stimuli presented were randomized across speaker variety. Each variety had the same probability of being selected on a given trial, such that, on average, a given participant heard each variety approximately eight times (See online supplementary materials for more information). Prior to pre-registering our research questions and hypotheses we piloted the 2AFC experiment on 100 monolingual Spanish speakers to assure the quality of the items and assess the difficulty of the task. We did not come across any issues.

**LexTALE**. To assess Spanish proficiency we administered the Lexical Test for Advanced Learners of Spanish (LexTALE-ESP, henceforth LexTALE) (Izura, Cuetos, & Brysbaert, 2014; Lemhöfer & Broersma, 2012). The LexTALE is a lexical decision experiment used to provide a standardized assessment of proficiency/vocabulary size in Spanish. In this task participants see a series of words on the computer screen and must decide if they are real or fake using the keyboard (‘1’ for real, ‘0’ for fake). LexTALE scores can range from −20 to 60. Monolingual Spanish speakers generally score above 50. Scores from individuals with little or no knowledge of Spanish tend to be negative. Adult learners with low to medium proficiency can range from 0 to 25, and advanced learners generally score above 25. We conceive of proficiency as a continuous variable, thus we consider a monolingual English speaker to have little to no proficiency in Spanish (i.e., a negative value on the LexTALE). In our data set participant scores ranged from −16 to 55, suggesting all proficiency levels were likely represented in the sample. The mean score was 12.95 (95% CrI: [11.18, 14.72]) with a standard deviation of 13.60 (95% CrI: [12.38, 14.9].

**Empathy Questionnaire**. The construct empathy was assessed using the Empathy Quotient (EQ, Baron-Cohen & Wheelwright, 2004). The EQ is a 60-item questionnaire that presents four point likert-type items ranging from ‘strongly agree’ to ‘strongly disagree’. Forty of the questions assess empathy and 20 are filler items. In order to avoid response bias, choices indicating empathic responses are coded to elicit “agree” responses in half the target items and “disagree” responses in the other half. The target items are scored with 2 or 1 points based on if the participant responds “strongly” or “slightly”. Finally, the EQ is scored by summing the total points to produce a single value indicating an individual’s level of empathy. Thus the minimum possible value is 0 (low empathy) and the maximum is 80 (high empathy). In our data set the average empathy quotient was 37.88 (Range: [9, 69], 95% CrI: [36.13, 39.68], SD: 13.39, 95% CrI of SD: [12.28, 14.67]). The empathy quotient in its entirety is available in the supplementary materials.

## Procedure

Participants recruited via Prolific.ac completed the all three tasks in a single session. The 2AFC task was first, followed by the LexTALE task, and, finally, the empathy quotient questionnaire. We planed to collect data from approximately 300 individuals: 100 monolingual Spanish speakers not reported here and 200 L2 learners). Following Brandl et al. (2020), we assumed the effect size for perceptual learning was moderate in terms of the criteria set forth for L2 research by Plonsky and Oswald (2014) (Cohen’s D = 0.600, Pearson’s r = 0.287). Based on this assumption, we estimated that we would need 94 participants to have an 80% chance of capturing the proficiency effect with a type II error rate of 5%. Our hypothesis related to empathy as a possible mediator of perceptual learning is exploratory in nature and thus we did not base our sample size estimate on any parameter estimates related to this effect. That said, we believed the aforementioned exploratory effect was likely to be small, and, considering the resources necessary and available to us, planned to recruit 100 additional participants.

We excluded data from participants in the following circumstances: error during data collection, clear lack of understanding or engagement during the task (i.e., all ‘1’ responses, failed three attention checks, etc.), participants reporting having learned Spanish before the age of 13, or participants with knowledge of language other and English and Spanish. Data from a total of 78 participants was discarded because the experimental session timed out and/or data was incomplete. An additional 8 participants were discarded due to low accuracy (n = 5), incomplete data (n = 2), and failed attention checks (n = 1). A total of 225 participants met the criteria for inclusion.

## Statistical analyses

We report three primary statistical analyses that were pre-registered prior to collecting the learner data: response accuracy, response times, and drift diffusion models. All additional analyses are exploratory in nature and explicitly described as such. First, we analyzed response accuracy using Bayesian multilevel logistic regression. The model considered response accuracy for the population effects *utterance type* (declarative broad focus, declarative narrow focus, interrogative yes/no, interrogative -wh), *LexTALE score* (i.e., proficiency), *empathy quotient*, and the higher order interactions. The likelihood of the model was Bernoulli distributed with a logit link function. The criterion, *response*, was coded as ‘1’ for correct responses and ‘0’ for incorrect responses. Thus, the first analysis modeled the probability of responding correctly to the prompt “Is this a question?”. We specified group-level effects for participants, speaker variety, and items. The slope for *utterance type* varied for the participant effect, as did the *LexTALE* by *empathy quotient* interaction for the speaker variety effect. All continuous variables were standardized and ‘interrogative yes/no’ was set as the baseline for *utterance type*, thus the model intercept represented the probability of a learner with average proficiency and average empathy responding correctly to a yes/no question.

The same model was fit to the response time data with the exception of the model likelihood, which was assumed to be distributed as lognormal. We also fit an additional exploratory model with the same population- and grouping-effects structure using d’ (d prime) as the outcome variable.

The third analysis utilized Bayesian drift diffusion modeling (DDM, Ratcliff & McKoon, 2008). This approach to analyzing behavioral data models decision-making as a random-walk decision process. DDMs can simultaneously take into account responses and response times in two-choice tasks in a single model, thus they are particularly beneficial when analyzing tasks in which speed-accuracy tradeoffs may be present. We estimate the parameters off the DDM using a Bayesian methods and subsequently fit measurement error models on the posterior estimates of the resulting parameters.

A DDM estimates four parameters: boundary separation, bias, drift rate and non-decision time. Boundary separation, α, quantifies the amount of information necessary to make a decision. The boundaries represent the thresholds for the two alternatives in the task, which, in our case, implies correct and incorrect responses. Bias, β, gives an indication of a preference for one of the choices at the beginning of the decision making process. A positive bias value indicates a preference for the upper boundary, whereas a negative bias is an indicator of a preference for the lower boundary. The drift rate, δ, provides an assessment of the rate at which information is accumulated. A higher δ implies a random walk that arrives at one of the thresholds faster and is interpreted as indication that the participant finds the task to be easier. Conversely a lower drift rate is interpreted as indicating a more difficult task. The sign of the value is also relevant. Positive drift rate refers to evidence accumulation for the upper boundary and negative drift rate for the lower boundary. Finally, non-decision time, τ, models the part of the time course that is not associated with decision making (e.g., the time necessary to perceive a stimuli prior to evidence accumulation). Figure 1 provides an example of a hypothetical DDM for the 2AFC task in the present project.

![Figure 1.   A drift diffusion model of the present study. The upper and lower bounds represent correct and incorrect responses, respectively. The boundary separation (α) is the distance between the two thresholds and indicates the evidence required to make a decision. Non-decision time (τ) represents the time course before evidence accumulation begins, i.e., time used for any process except decision making. Bias (β) is the starting point for the evidence accumulation in the vertical plane (i.e., closer or further away from a given threshold), and drift rate (δ) quantifies the rate of evidence accumulation. The purple and orange lines represent examples of a decision resulting in a correct (purple) and incorrect (orange) decision. The corresponding density curves represent the distribution of response times at either threshold.](data:application/pdf;base64,)

*Figure* *1.*  A drift diffusion model of the present study. The upper and lower bounds represent correct and incorrect responses, respectively. The boundary separation (α) is the distance between the two thresholds and indicates the evidence required to make a decision. Non-decision time (τ) represents the time course before evidence accumulation begins, i.e., time used for any process except decision making. Bias (β) is the starting point for the evidence accumulation in the vertical plane (i.e., closer or further away from a given threshold), and drift rate (δ) quantifies the rate of evidence accumulation. The purple and orange lines represent examples of a decision resulting in a correct (purple) and incorrect (orange) decision. The corresponding density curves represent the distribution of response times at either threshold.

We estimated the aforementioned parameters by fitting a DDM to the response and response time data of each participant independently. We opted for this approach, as opposed to fitting a single model including all participants, for computational reasons. Put simply, the model likely would have taken weeks to fit, whereas the no-pooling (i.e., by-participant) method took approximately 26 hours. Thus after fitting the DDMs, we obtained a posterior distribution of plausible values for boundary separation, drift rate, bias, and non-decision time for each participant. Next, we used measurement-error models to analyze boundary separation (α) and drift rate (δ) independently. These models followed the same functional form as the response accuracy model described above. That is, in two separate models, we analysed the boundary separation and drift rate data as a function of *utterance type* (interrogative yes/no, interrogative -wh), *LexTALE score* (i.e., proficiency), *empathy quotient*, and the higher order interactions. The primary difference between the measurement-error models and the traditional regression analyses described for the response data is that the former can incorporate a measure of uncertainty around a point estimate. To give a concrete example, the analysis of the boundary separation data included the posterior median and the standard error for each participant as the outcome variable, as opposed to using just a single point estimate.

For all models, we included regularizing, weakly informative priors (Gelman, Simpson, & Betancourt, 2017). Generally, we sample the posterior distribution of a given model for statistical inferences. To assess our pre-registered hypotheses we established a region of practical equivalence (ROPE) around a point null value of 0 (see Kruschke, 2018) using the following formula:

For all models, median posterior point estimates are reported for each parameter of interest, along with the 95% highest density interval (HDI), the percent of the region of the HDI contained within the ROPE, and the maximum probability of effect (MPE). For statistical inferences, a posterior distribution for a parameter β in which 95% of the HDI falls outside the ROPE and a high MPE (i.e., values close to 1) were taken as compelling evidence for a given effect. All exploratory analyses, explicitly described as such, include posterior point estimates, the 95% HDI, and the MPE. We conducted all analyses using R and fit all models using the probabilistic programming language stan via the R package brms (Bürkner, 2017, 2018). Finally, we provide more information for all analyses in the supplementary materials.

# Results

## Response accuracy and response times

Figure 2 (left panel) summarizes the posterior distribution of the omnibus response accuracy model, illustrating point estimates ±66% and 95% HDIs in graphical form. An equivalent summary of the posterior distribution in table format is provided in the supplementary materials (See Table 1. The log-odds of a correct response to an absolute interrogative utterance at the average proficiency and EQ levels was 0.53, or approximately 62.95% (β = 0.53, HDI = [0.23, 0.83], ROPE = 0, MPE = 1). In comparison, all other utterance types were associated with an increase in the log odds of responding correctly. The right panel of Figure 2 plots response accuracy of each utterance type in the probability space. As illustrated in the plot, participants were slightly more accurate when responding to wh- questions (β = 0.43, HDI = [0.17, 0.65], ROPE = 0, MPE = 1) with approximately 72.31% correct, and much more accurate when responding to declarative utterances (narrow focus: β = 2.13, HDI = [1.86, 2.40], ROPE = 0, MPE = 1, Accuracy = 93.46%; broad focus: β = 2.34, HDI = [2.06, 2.60], ROPE = 0, MPE = 1, Accuracy = 94.63%).[[2]](#footnote-31)

![Figure 2.   Forest plot summary of the response accuracy model (left panel) and posterior probability of a correct response for each utterance type (right panel). For both plots, white points represent posterior medians ±66% and 95% highest density credible intervals.](data:application/pdf;base64,)

*Figure* *2.*  Forest plot summary of the response accuracy model (left panel) and posterior probability of a correct response for each utterance type (right panel). For both plots, white points represent posterior medians ±66% and 95% highest density credible intervals.

Figure 3 plots response accuracy as a function of utterance type and proficiency (left panel) and empathy quotient (right panel). For all utterance types response accuracy increased as proficiency increased. Though the proficiency effect was most visually obvious for y/n questions (β = 0.28, HDI = [0.15, 0.41], ROPE = 0, MPE = 1) and wh- questions (β = 0.40, HDI = [0.24, 0.57], ROPE = 0.00, MPE = 1.00), this was also the case for broad focus (β = 0.48, HDI = [0.26, 0.71], ROPE = 0.00, MPE = 1.00) and narrow focus (β = 0.31, HDI = [0.10, 0.51], ROPE = 0.00, MPE = 1.00) declaratives. There was no evidence that empathy level predicted response accuracy for y/n questions (β = −0.02, HDI = [−0.12, 0.09], ROPE = 0.98, MPE = 0.62), however for wh- questions (β = 0.18, HDI = [0.05, 0.32], ROPE = 0.10, MPE = 1.00), broad focus declaratives (β = 0.23, HDI = [0.05, 0.43], ROPE = 0.06, MPE = 0.99), and narrow focus (β = 0.24, HDI = [0.07, 0.42], ROPE = 0.03, MPE = 1.00) declaratives we find compelling evidence that the effect is positive.

![Figure 3.   Response accuracy: Probability of a correct response in the logistic space as a function of LexTALE score (left panel) and empathy quotient (right panel) for each utterance type. Thin lines represent 300 draws from the posterior distribution for each condition.](data:application/pdf;base64,)

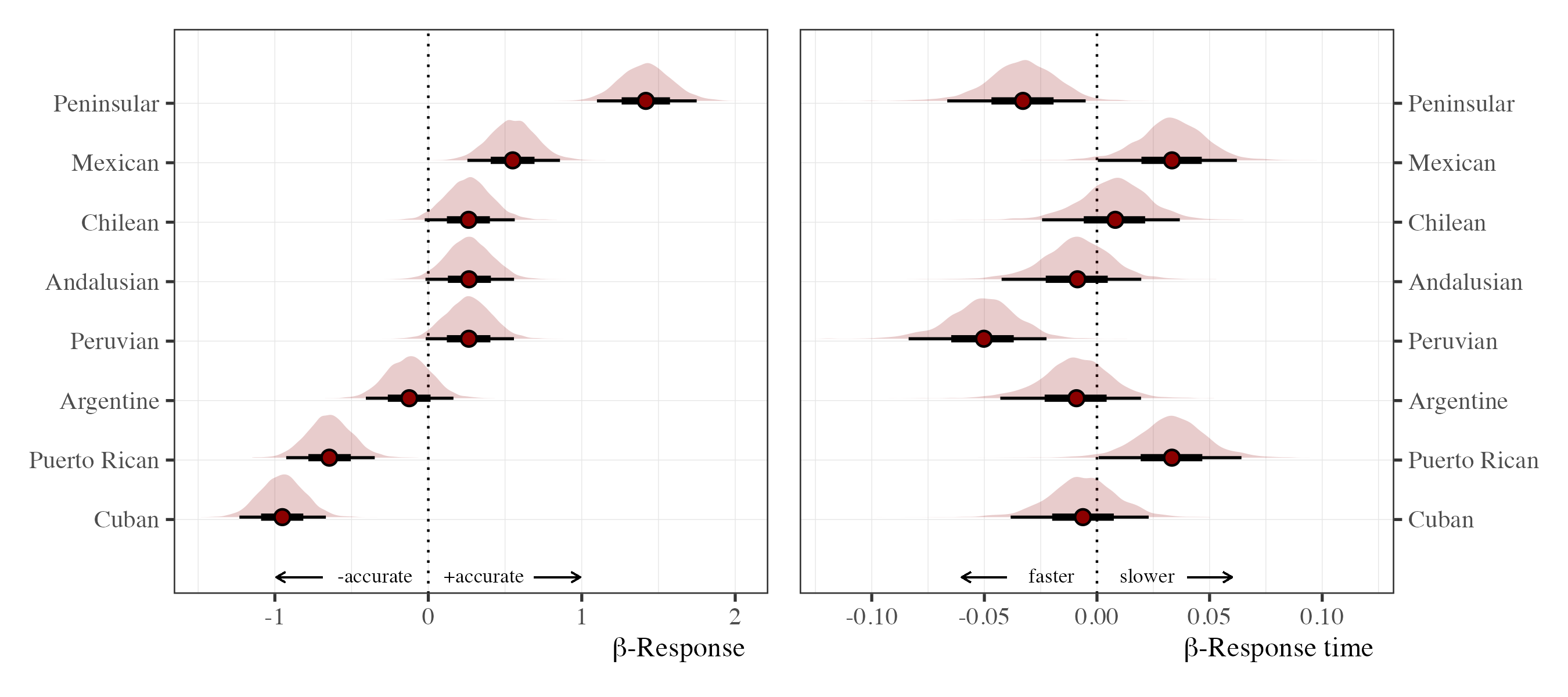
*Figure* *3.*  Response accuracy: Probability of a correct response in the logistic space as a function of LexTALE score (left panel) and empathy quotient (right panel) for each utterance type. Thin lines represent 300 draws from the posterior distribution for each condition.

The omnibus model also estimated the proficiency × empathy quotient simple interaction for each utterance type. We used the posterior distribution to estimate the probability that this effect was non-zero for each utterance type. We found evidence that the proficiency effect was modulated by empathy quotient scores for wh- questions (β = 0.22, HDI = [0.05, 0.39], ROPE = 0.06, MPE = 0.99), though not for y/n questions (β = 0.02, HDI = [−0.09, 0.14], ROPE = 0.92, MPE = 0.65), broad focus declaratives (β = 0.10, HDI = [−0.14, 0.35], ROPE = 0.46, MPE = 0.80), nor narrow focus declaratives (β = 0.04, HDI = [−0.17, 0.25], ROPE = 0.65, MPE = 0.64). This relationship is illustrated in Figure 4. Specifically, we plot conditional effects of response accuracy as a function of proficiency and empathy quotient for the y/n and wh- interrogatives. In the left panel of Figure 4, one observes a positive correlation between response accuracy and proficiency that remains constant at standardized empathy quotient values of -1, 0 and +1 for the y/n questions. For the wh- questions (right panel), on the other hand, we see that the slope of the proficiency effect increases for higher empathy quotient values.  
That is to say, for wh- questions, at a given proficiency level, learners with higher empathy (black lines) tended to respond more accurately.

![Figure 4.   Response accuracy: Probability of a correct response as a function of empathy quotient and LexTALE score for each question type.](data:application/pdf;base64,)

*Figure* *4.*  Response accuracy: Probability of a correct response as a function of empathy quotient and LexTALE score for each question type.

The speaker variety grouping effect from the omnibus model provided estimates of overall participant accuracy for each speaker variety (See Figure 5). As was the case with the monolingual Spanish pilot data, learners were least accurate when responding to the Cuban variety and most accurate when responding to the standard Peninsular and Mexican varieties. Response accuracy to a given variety did not correlate with response times. For instance, as mentioned, learners were least accurate when responding to the Cuban stimuli, but had average response times similar to the grand mean for this variety.



*Figure* *5.*  Group-level estimates of response accuracy and response time as a function of speaker variety. Red points represent posterior medians ±66% and 95% highest density credible intervals. The vertical dotted lines indicate the grand mean.

## Drift diffusion models

This is Figure 6

![Figure 6.   Need a caption.](data:application/pdf;base64,)

*Figure* *6.*  Need a caption.

This is Figure 7

![Figure 7.   Two-thousand simulations of the drift diffusion model for interrogative utterances as a function of empathy quotient (low/high) and LexTALE score (low/high). Low and high levels represent ±2 standard deviations above/below the mean. Dark red lines represent the average decision time course.](data:application/pdf;base64,)

*Figure* *7.*  Two-thousand simulations of the drift diffusion model for interrogative utterances as a function of empathy quotient (low/high) and LexTALE score (low/high). Low and high levels represent ±2 standard deviations above/below the mean. Dark red lines represent the average decision time course.

# Discussion

# Conclusion

# References

Astruc, L., Vanrell, M. del M., & Prieto, P. (2016). Cost of the action and social distance affect the selection of question intonation in Catalan. In M. E. Armstrong, N. Henriksen, & M. del M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 93–113). John Benjamins Publishing Company. <https://doi.org/10.1075/ihll.6>

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

Baron-Cohen, S. (2011). *Zero degree of empathy. On empathy and the origins of cruelty*. London, England: Penguin.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

Bishop, J., & Kuo, G. (2016). Do “autistic-like” personality traits predict prosody perception? *Talk presented at LabPhon15 satellite workshop on personality in speech perception and production, Ithaca, NY.*

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [computer program]*. Retrieved from <http://www.praat.org/>

Brandl, A., González, C., & Bustin, A. (2020). The development of intonation in L2 Spanish: A perceptual study. In A. Morales-Front, M. J. Ferreira, R. P. Leow, & C. Sanz (Eds.), *Hispanic linguistics: Current issues and new directions* (pp. 12–31). John Benjamins Publishing Company. <https://doi.org/10.1075/ihll.26>

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–681. <https://doi.org/10.1037/0022-0663.96.4.671>

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*(2), 121–138. <https://doi.org/10.1017/S0140525X09000545>

Chen, A. (2005). *On the universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT.

Craft, J. (2015). *The acquisition of intonation by L2 Spanish speakers while on a six week study abroad program in Valencia, Spain* (PhD thesis). The Florida State University.

Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Berlin: Mouton de Gruyter.

Degen, J., & Tanenhaus, M. K. (2016). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, *40*(1), 172–201. <https://doi.org/10.1111/cogs.12227>

de-la-Mota, C. (2019). Improving non-native pronunciation: Teaching prosody to learners of Spanish as a second/foreign language. In R. Rao (Ed.), *Key issues in the teaching of Spanish pronunciation* (pp. 162–197). Routledge.

Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B., & D’Imperio, M. (2016). *The role of individual empathic skills on the online processing of intonational meaning*. Bilbao, Spain: Basque Center on Cognition, Brain; Language. <https://doi.org/10.13140/RG.2.2.19401.13926>

Esteve-Gibert, N., Schafer, A. J., Hemforth, B., Portes, C., Pozniak, C., & D’Imperio, M. (2020). Empathy influences how listeners interpret intonation and meaning when words are ambiguous. *Memory & Cognition*, *48*, 566–580. <https://doi.org/10.3758/s13421-019-00990-w>

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901.002>

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 459–473. <https://doi.org/10.1098/rstb.2002.1218>

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 1–13. <https://doi.org/10.3390/e19100555>

Henriksen, N., Armstrong, M. E., & Garcı́a-Amaya, L. (2016). The intonational meaning of polar questions in Manchego Spanish spontaneous speech. In M. E. Armstrong, N. Henriksen, & M. del M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 181–205). John Benjamins Publishing Company. <https://doi.org/10.1075/ihll.6>

Henriksen, N., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone Linguistics*, *3*(1), 113–162. https://doi.org/<https://doi.org/10.1515/shll-2010-1067>

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, *127*(3), 366–376. <https://doi.org/10.1016/j.bandl.2012.11.006>

Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota & P. Prieto (Eds.), *Intonation in romance* (pp. 350–391). Oxford University Press.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). LexTALE-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49–66. <https://doi.org/10.1037/t47086-000>

Jilka, M. (2000). Testing the contribution of prosody to the perception of foreign accent. In A. James & J. Leather (Eds.), *In proceedings of new sounds 4th international symposium on the acquisition of second language speech. Amsterdam: University of amsterdam* (Vol. 4, pp. 199–207).

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280.

Kvavik, K. H., & Olsen, C. L. (1974). Theories and methods in spanish intonational studies. *Phonetica*, *30*(2), 65–100. <https://doi.org/10.1159/000259481>

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>

Levis, J. (2016). Accent in second language pronunciation research and teaching. *Journal of Second Language Pronunciation*, *2*(2), 153–159. <https://doi.org/10.1075/jslp.2.2.01lev>

Liu, Y. (2017). Study on the influence of emotion factors in Second Language Acquisition. In G. Yu, G. Ke, & L. Han (Eds.), *Proceedings of the International Conference on Financial Management, Education and Social Science* (pp. 261–264). <https://doi.org/10.25236/fmess.2017.55>

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody. Phonetic description and teaching practice* (pp. 53–76). Berlin: De Gruyter Mouton.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, *17*(1), 17–34. <https://doi.org/10.1017/S0272263100013735>

Nibert, H. J. (2005). The acquisition of the phrase accent by intermediate and advanced adult learners of Spanish as a second language. *Selected proceedings of the 6th conference on the acquisition of spanish and portuguese as first and second languages*, 108–122. Somerville, MA: Cascadilla Proceedings Project.

Nibert, H. J. (2006). The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. *Selected proceedings of the 2nd conference on laboratory approaches to spanish phonetics and phonology*, 131–148. Somerville, MA: Cascadilla Proceedings Project.

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, *63*(3), 324–346. <https://doi.org/10.1016/j.jml.2010.06.005>

Orrico, R., & D’Imperio, M. (2020). Individual empathy levels affect gradual intonation-meaning mapping: The case of biased questions in salerno italian. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *11*(1). <https://doi.org/10.5334/labphon.238>

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., … Lindelv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pettorino, M., De Meo, A., & Vitale, M. (2014). Transplanting vowels towards the acoustic correlates of foreign accent. In Y. Congosto, M. L. Montero Curiel, & A. Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación: II. Adquisición y aprendizaje de lenguas/español como lengua extranjera* (pp. 93–106).

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *Tesol Quarterly*, *35*(2), 233–255. <https://doi.org/10.2307/3587647>

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. <https://doi.org/10.1111/lang.12079>

Rao, R. (2019). Introduction. In R. Rao (Ed.), *Key issues in the teaching of Spanish pronunciation* (pp. 1–13). Routledge.

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

Rota, G., & Reiterer, S. M. (2009). Cognitive aspects of pronunciation talent. In G. Dogil & S. M. Reiterer (Eds.), *Language talent and brain activity* (pp. 67–96). Mouton de Gruyter. <https://doi.org/10.1515/9783110215496>

Thornberry, P. A. (2014). *The L2 acquisition of Buenos Aires Spanish intonation during a study abroad semester* (PhD thesis). University of Minnesota.

Trimble, J. C. (2013a). *Acquiring variable L2 spanish intonation in a study abroad context* (PhD thesis). University of Minnesota.

Trimble, J. C. (2013b). Perceiving intonational cues in a foreign language: Perception of sentence type in two dialects of Spanish. In C. Howe (Ed.), *Selected proceedings of the 15th hispanic linguistics symposium* (pp. 78–92). Somerville, MA: Cascadilla Proceedings Project.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. <https://doi.org/10.1017/S0272263106060013>

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 6–12. <https://doi.org/10.3389/fpsyg.2015.01000>

Appendix

# On-line supplementary material

## Traditional analyses

This section contains additional information regarding the response accuracy and response time analyses, as well as tables reported but not included in the main text.

**Learner response accuracy**. The population effects of the response accuracy model were specified in the following manner:

We employed the 0 + Intercept syntax of brms and set weakly informative priors as follows:

The summary of the response accuracy model is avaibale in Table 1. The information provided in this table is equivalent to the left panel of Figure 2 in the manuscript.

Table 1: Summary of the posterior distribution modeling response accuracy as a function of utterance type, LexTALE, and Empathy quotient. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

| Parameter | Median | HDI | % in ROPE | MPE | Rhat | ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 0.53 | [0.23, 0.83] | 0.00 | 1.00 | 1.00 | 2099 |
| Int. wh- | 0.43 | [0.17, 0.65] | 0.00 | 1.00 | 1.00 | 2723 |
| Dec. narrow focus | 2.13 | [1.86, 2.40] | 0.00 | 1.00 | 1.00 | 2608 |
| Dec. broad focus | 2.34 | [2.06, 2.60] | 0.00 | 1.00 | 1.00 | 2576 |
| LexTALE | 0.28 | [0.15, 0.41] | 0.00 | 1.00 | 1.00 | 5245 |
| Empathy quotient | −0.02 | [−0.12, 0.09] | 0.98 | 0.62 | 1.00 | 5351 |
| Int. wh-:LexTALE | 0.12 | [−0.05, 0.30] | 0.42 | 0.90 | 1.00 | 4977 |
| Dec. narrow focus:LexTALE | 0.02 | [−0.17, 0.22] | 0.72 | 0.58 | 1.00 | 8450 |
| Dec. broad focus:LexTALE | 0.19 | [−0.02, 0.42] | 0.18 | 0.96 | 1.00 | 9482 |
| Int. wh-:EQ | 0.20 | [0.03, 0.36] | 0.10 | 0.99 | 1.00 | 4912 |
| Dec. narrow focus:EQ | 0.26 | [0.08, 0.43] | 0.02 | 1.00 | 1.00 | 8389 |
| Dec. broad focus:EQ | 0.24 | [0.05, 0.43] | 0.05 | 0.99 | 1.00 | 8180 |
| LexTALE:EQ | 0.02 | [−0.09, 0.14] | 0.92 | 0.65 | 1.00 | 5954 |
| Int. wh-:LexTALE:EQ | 0.19 | [0.00, 0.40] | 0.16 | 0.97 | 1.00 | 5604 |
| Dec. narrow focus:LexTALE:EQ | 0.02 | [−0.20, 0.23] | 0.67 | 0.56 | 1.00 | 8344 |
| Dec. broad focus:LexTALE:EQ | 0.08 | [−0.16, 0.33] | 0.51 | 0.74 | 1.00 | 8698 |

## Drift diffusion models

Drift Diffusion Models (DDM), also referred to as Wiener Diffusion Models and Decision Diffusion Models, represent our preferred method for analyzing the data from our 2AFC task. DDMs are rarely used in SLA research, though they are commonplace in psychology. The primary selling point of using a DDM is related to the parameters the model estimates: boundary separation (α), drift rate (δ), bias (β), and non-decision time (τ). Together these parameters give rich information about the processes believed to underpin decision making. Specifically, a DDM requires decision data, e.g., “left” or “right” choices, correct or incorrect responses, etc., and response times associated with said decisions. In linguistics, particularly in psycholinguistics, data of this nature derived from 2AFC tasks are often analyzed using separate models, one for responses, and another for response times (as we have done in our so-called ‘traditional analyses’).[[3]](#footnote-128) As mentioned, a DDM uses both of these dependent variables—responses and repsonse times—to estimate the 4 aforementioned parameters. The estimates can then be scrutinized in subsequent models, if one estimates the parameters for each participant (i.e., the approach taken in the present work), and/or used for simulations. For our purposes, we employ the Bayesian implementation of the DDM, thus we sample from a posterior distribution of plausible estimates for α, δ, β, and τ for each participant. We then summarize and report these posterior distributions for statistical inferences.

The no-pooling models were fit using the following specification in brms:

rt\_raw | dec(is\_correct) ~ 0 + sentence\_type,  
 bs ~ 0 + sentence\_type,   
 ndt ~ 0 + sentence\_type,   
 bias ~ 0 + sentence\_type

and the priors were:

prior("normal(0, 1)", class = "b"),  
prior("normal(0, 5)", class = "b", dpar = "bs"),  
prior("normal(0.2, 1)", class = "b", dpar = "ndt"),  
prior("normal(0.5, 1)", class = "b", dpar = "bias")

The complete code used to fit the models are available in 09\_ddm.R in the r scripts directory.

**Measurement-error models**. The measurement error models fit to the boundary separation and drift rate data were specified to include the standard error around each posterior median for α and δ:

The priors for the drift rate model were:

and the priors for the boundary separation model were:

To specify this type of model in brms we use the resp\_se function, as follows:

estimate | resp\_se(se, sigma = TRUE) ~ 1 + # Criterion  
 q\_sum \* lextale\_std \* eq\_std + # Population-level effects  
 (1 + q\_sum \* lextale\_std \* eq\_std | participant) # Group-level effects

Table 2 is equivalent to Figure 6 in the main document.

Table 2: Summary of the posterior distribution modeling boundary separation and drift rate as a function of question type, LexTALE, and Empathy quotient. The table includes posterior medians, the 95% HDI, the percentage of the HDI within the ROPE, and the maximum probability of effect (MPE).

| Model | Parameter | Median | HDI | MPE | Rhat | ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Boundary | Intercept | 1.77 | [1.70, 1.83] | 1.00 | 1.00 | 3407 |
| separation | Question type | −0.04 | [−0.08, −0.01] | 0.99 | 1.00 | 3676 |
|  | LexTALE | 0.14 | [0.06, 0.22] | 1.00 | 1.00 | 3460 |
|  | EQ | 0.04 | [−0.02, 0.11] | 0.91 | 1.00 | 3585 |
|  | Question type:LexTALE | −0.05 | [−0.09, 0.00] | 0.97 | 1.00 | 3594 |
|  | Question type:EQ | −0.01 | [−0.04, 0.03] | 0.73 | 1.00 | 3993 |
|  | LexTALE:EQ | 0.12 | [0.03, 0.20] | 1.00 | 1.00 | 3912 |
|  | Question type:LexTALE:EQ | −0.02 | [−0.07, 0.03] | 0.77 | 1.00 | 3580 |
| Drift rate | Intercept | 1.23 | [1.20, 1.26] | 1.00 | 1.00 | 3814 |
|  | Question type | 0.08 | [0.06, 0.10] | 1.00 | 1.00 | 3584 |
|  | LexTALE | 0.02 | [−0.02, 0.05] | 0.83 | 1.00 | 3276 |
|  | EQ | 0.00 | [−0.03, 0.03] | 0.59 | 1.00 | 4063 |
|  | Question type:LexTALE | 0.01 | [−0.02, 0.05] | 0.70 | 1.00 | 3846 |
|  | Question type:EQ | 0.00 | [−0.02, 0.02] | 0.53 | 1.00 | 4123 |
|  | LexTALE:EQ | −0.06 | [−0.11, −0.02] | 1.00 | 1.00 | 4114 |
|  | Question type:LexTALE:EQ | 0.01 | [−0.03, 0.05] | 0.66 | 1.00 | 3733 |

## Supplementary analyses

In this section we present supplementary analyses, all of which are exploratory in nature.

**D’**. Figure 8 and Table 3 represent an exploratory analysis of d’ scores as a function of utterance type and speaker variety. One observes similar patterns to those from the accuracy analysis presented in the manuscript. The primary takeaway is that the analysis of learners’ sensitivity to Spanish prosody mirrors that their accuracy. That is to say, learners are more sensitive to (and accurate with) statements (declarative broad focus, declarative narrow focus) than questions (interrogative wh-, interrogative y/n) (left panel of Figure 8. Learner sensitivity to the different Spanish varieties represented in the stimuli pattern in the same manner, i.e., more sensitivity to the Peninsular variety and less sensitivity to the Cuban variety (right panel of Figure 8). Table 3 summarizes the posterior of these exploratory analyses.

![Figure 8.   Exploratory analysis of d’ as a function of utterance type and speaker variety. Points represent posterior medians ±66% and 95% credible intervals.](data:application/pdf;base64,)

*Figure* *8.*  Exploratory analysis of d’ as a function of utterance type and speaker variety. Points represent posterior medians ±66% and 95% credible intervals.

Table 3: Summary of the posterior distribution modeling d’ as a function of question type or speaker variety. The table includes posterior medians, the 95% HDI, and the maximum probability of effect (MPE).

| Model | Parameter | Median | HDI | MPE |
| --- | --- | --- | --- | --- |
| Utterance type | Declarative broad focus | 1.18 | [1.14, 1.23] | 1.00 |
|  | Declarative narrow focus | 1.14 | [1.10, 1.19] | 1.00 |
|  | Interrogative wh- | 0.54 | [0.47, 0.61] | 1.00 |
|  | Interrogative y/n | 0.25 | [0.19, 0.31] | 1.00 |
| Variety | Andalusian | 1.48 | [1.38, 1.59] | 1.00 |
|  | Argentine | 1.32 | [1.22, 1.42] | 1.00 |
|  | Chilean | 1.49 | [1.40, 1.59] | 1.00 |
|  | Cuban | 0.76 | [0.67, 0.86] | 1.00 |
|  | Mexican | 1.70 | [1.61, 1.80] | 1.00 |
|  | Peninsular | 2.07 | [1.98, 2.15] | 1.00 |
|  | Peruvian | 1.51 | [1.42, 1.61] | 1.00 |
|  | Puerto Rican | 0.95 | [0.85, 1.05] | 1.00 |

**Randomization check across participants**. For the purposes of our research questions, it was important that every participant be presented with stimuli from all of the Spanish varieties to which we had access. Recall that the 2AFC task contained 64 items, 16 of each utterance type. Using javascript we assigned each variety an equal probability of being selected in a given trial (0.125). To ensure that our randomization worked as planned (i.e., with each variety represented approximately equally across all trials and all participants), we calculated the average number of times each variety was presented in the data set (n = 225, and 14400 trials). One can observe in Figure 9 that this is indeed the case.

![Figure 9.   Make sure each variety represented equally over course of experiment.](data:application/pdf;base64,)

*Figure* *9.*  Make sure each variety represented equally over course of experiment.

*Speech rate*. This is Figure 10

![Figure 10.   Make sure each variety represented equally over course of experiment.](data:application/pdf;base64,)

*Figure* *10.*  Make sure each variety represented equally over course of experiment.

## Reproducibility information

**About this document**

This document was written in RMarkdown using papaja (Aust & Barth, 2018).

**Session info**

setting value  
 version R version 4.1.0 (2021-05-18)  
 os macOS Big Sur 10.16  
 system x86\_64, darwin17.0  
 ui X11  
 language (EN)  
 collate en\_US.UTF-8  
 ctype en\_US.UTF-8  
 tz America/New\_York  
 date 2022-01-19  
 pandoc 2.14.2 @ /Applications/RStudio.app/Contents/MacOS/pandoc/ (via rmarkdown)

loadedversion date  
abind 1.4-5 2016-07-21  
arrayhelpers 1.1-0 2020-02-04  
assertthat 0.2.1 2019-03-21  
backports 1.4.1 2021-12-13  
base64enc 0.1-3 2015-07-28  
bayesplot 1.8.1 2021-06-14  
bayestestR 0.11.5 2021-10-30  
beeswarm 0.4.0 2021-06-01  
bit 4.0.4 2020-08-04  
bit64 4.0.5 2020-08-30  
bookdown 0.24 2021-09-02  
boot 1.3-28 2021-05-03  
bridgesampling 1.1-2 2021-04-16  
brms 2.16.3 2021-11-22  
Brobdingnag 1.2-6 2018-08-13  
cachem 1.0.6 2021-08-19  
callr 3.7.0 2021-04-20  
cellranger 1.1.0 2016-07-27  
checkmate 2.0.0 2020-02-06  
cli 3.1.0 2021-10-27  
cmdstanr 0.4.0 2021-07-22  
coda 0.19-4 2020-09-30  
codetools 0.2-18 2020-11-04  
colorspace 2.0-2 2021-06-24  
colourpicker 1.1.1 2021-10-04  
contributoR 0.3.0 2021-12-18  
crayon 1.4.2 2021-10-29  
crosstalk 1.2.0 2021-11-04  
curl 4.3.2 2021-06-23  
datawizard 0.2.1 2021-10-04  
DBI 1.1.1 2021-01-15  
desc 1.4.0 2021-09-28  
devtools 2.4.3 2021-11-30  
digest 0.6.29 2021-12-01  
distributional 0.2.2 2021-02-02  
dplyr 1.0.7 2021-06-18  
DT 0.20 2021-11-15  
dygraphs 1.1.1.6 2018-07-11  
ellipsis 0.3.2 2021-04-29  
emmeans 1.7.1-1 2021-11-29  
estimability 1.3 2018-02-11  
evaluate 0.14 2019-05-28  
extrafont 0.17 2014-12-08  
extrafontdb 1.0 2012-06-11  
fansi 0.5.0 2021-05-25  
farver 2.1.0 2021-02-28  
fastmap 1.1.0 2021-01-25  
forcats 0.5.1 2021-01-27  
fs 1.5.2 2021-12-08  
gamm4 0.2-6 2020-04-03  
ganttrify 0.0.0.9007 2021-07-19  
generics 0.1.1 2021-10-25  
ggbeeswarm 0.6.0 2017-08-07  
ggdist 3.0.1 2021-11-30  
ggplot2 3.3.5 2021-06-25  
ggrepel 0.9.1 2021-01-15  
ggridges 0.5.3 2021-01-08  
glue 1.5.1 2021-11-30  
gridExtra 2.3 2017-09-09  
gtable 0.3.0 2019-03-25  
gtools 3.9.2 2021-06-06  
here 1.0.1 2020-12-13  
highr 0.9 2021-04-16  
hms 1.1.1 2021-09-26  
htmltools 0.5.2 2021-08-25  
htmlwidgets 1.5.4 2021-09-08  
httpuv 1.6.3 2021-09-09  
igraph 1.2.9 2021-11-23  
inline 0.3.19 2021-05-31  
insight 0.14.5 2021-10-16  
jsonlite 1.7.2 2020-12-09  
knitr 1.36 2021-09-29  
later 1.3.0 2021-08-18  
lattice 0.20-45 2021-09-22  
lifecycle 1.0.1 2021-09-24  
lme4 1.1-27.1 2021-06-22  
loo 2.4.1 2020-12-09  
magrittr 2.0.1 2020-11-17  
markdown 1.1 2019-08-07  
MASS 7.3-54 2021-05-03  
Matrix 1.4-0 2021-12-08  
matrixStats 0.61.0 2021-09-17  
memoise 2.0.1 2021-11-26  
mgcv 1.8-38 2021-10-06  
mime 0.12 2021-09-28  
miniUI 0.1.1.1 2018-05-18  
minqa 1.2.4 2014-10-09  
munsell 0.5.0 2018-06-12  
mvtnorm 1.1-3 2021-10-08  
nlme 3.1-153 2021-09-07  
nloptr 1.2.2.3 2021-11-02  
papaja 0.1.0.9997 2021-12-11  
patchwork 1.1.1 2020-12-17  
pillar 1.6.4 2021-10-18  
pkgbuild 1.3.0 2021-12-09  
pkgconfig 2.0.3 2019-09-22  
pkgload 1.2.4 2021-11-30  
plyr 1.8.6 2020-03-03  
png 0.1-7 2013-12-03  
posterior 1.1.0 2021-09-09  
prettyunits 1.1.1 2020-01-24  
printy 0.0.0.9003 2021-09-26  
processx 3.5.2 2021-04-30  
projpred 2.0.2 2020-10-28  
promises 1.2.0.1 2021-02-11  
ps 1.6.0 2021-02-28  
purrr 0.3.4 2020-04-17  
R6 2.5.1 2021-08-19  
Rcpp 1.0.7 2021-07-07  
RcppParallel 5.1.4 2021-05-04  
readr 2.1.1 2021-11-30  
readxl 1.3.1 2019-03-13  
remotes 2.4.2 2021-11-30  
reshape2 1.4.4 2020-04-09  
rlang 0.4.12 2021-10-18  
rmarkdown 2.11 2021-09-14  
rprojroot 2.0.2 2020-11-15  
rsconnect 0.8.25 2021-11-19  
rstan 2.26.4 2021-10-18  
rstantools 2.1.1 2020-07-06  
rstudioapi 0.13 2020-11-12  
Rttf2pt1 1.3.9 2021-07-22  
scales 1.1.1 2020-05-11  
sessioninfo 1.2.2 2021-12-06  
shiny 1.7.1 2021-10-02  
shinyjs 2.0.0 2020-09-09  
shinystan 2.5.0 2018-05-01  
shinythemes 1.2.0 2021-01-25  
StanHeaders 2.26.4 2021-10-18  
stringi 1.7.6 2021-11-29  
stringr 1.4.0 2019-02-10  
svUnit 1.0.6 2021-04-19  
tensorA 0.36.2 2020-11-19  
testthat 3.1.1 2021-12-03  
threejs 0.3.3 2020-01-21  
tibble 3.1.6 2021-11-07  
tidybayes 3.0.1 2021-08-22  
tidyr 1.1.4 2021-09-27  
tidyselect 1.1.1 2021-04-30  
tinylabels 0.2.2 2021-12-06  
tzdb 0.2.0 2021-10-27  
usethis 2.1.5 2021-12-09  
utf8 1.2.2 2021-07-24  
V8 3.6.0 2021-11-10  
vctrs 0.3.8 2021-04-29  
vipor 0.4.5 2017-03-22  
vroom 1.5.7 2021-11-30  
withr 2.4.3 2021-11-30  
writexl 1.4.0 2021-04-20  
xfun 0.29 2021-12-14  
xtable 1.8-4 2019-04-21  
xts 0.12.1 2020-09-09  
yaml 2.2.1 2020-02-01  
zoo 1.8-9 2021-03-09

1. See <https://osf.io/dh4zp/?view_only=162d6d13e5814417bcb9de349f18cb62> [↑](#footnote-ref-22)
2. An exploratory (i.e., not pre-registered) analysis of sensitivity to utterance type was also conducted using d’ in lieu of response accuracy. The results mirrored those found in the response accuracy model. That is, participants showed highest sensitivity to the declarative utterances, followed by wh- questions and y/n questions. These exploratory analyses are reported in the supplementary materials (See Table 3 and Figure 8). [↑](#footnote-ref-31)
3. Given how relatively uncommon DDMs are in linguistics, the present work includes both approaches, though it is reasonable to assume that this practice will diminish as DDMs become more well-known and the resources for implementing become more user-friendly. [↑](#footnote-ref-128)