Anthony R Velasquez
L245B | Methods in Psycholinguistics
Replication Writeup
4 June 2021

Persona effects on semantic resolution: A replication study of Beltrama & Schwarz (under review)

1 Introduction

The relationship between form and meaning in language is not a simple one-to-one mapping which holds across all contexts of use and all speakers. Recent research has shown that social information of the speaker is processed early and automatically, and has an impact on phonological processes (D'Onofrio 2015, 2018) and syntactic alignment (Weatherholtz et al. 2014). Other work has linked semantic/pragmatic meaning to social meaning, showing that inferences about the latter can be made using the former (Acton & Potts 2014, Beltrama & Staum Casasanto 2017). However, the question of how social meaning impacts semantic/pragmatic meaning has only begun to be addressed in the literature.

A set of experiments carried out by Andrea Beltrama and Florian Schwarz (under review) showed that the identity of a speaker also impacts precision expectations in interpretation of numerical expressions of cost, time, and distance. Specifically, Beltrama and Schwarz presented participants with a pair of speakers who were visually stereotyped as Nerdy or Chill, with one of the characters asking a question and the other giving a response as a time, a distance, or a price, and asked participants to determine which of a pair of smartphones the speaker must have been looking at when they made their utterance. One phone screen faced the participant and displayed a number that matched, didn't match, or approximately matched the number uttered; the other phone was turned over, so that the screen wasn't visible. Their results show that participants are more likely to choose the covered-screen phone in approximate conditions when the speaker is Nerdy rather than Chill, suggesting that persona information impacts expectations about the numerical precision of a speaker, and therefore the interpretation of the utterance.

Here I report on an experiment conducted to replicate a part of Beltrama & Schwarz (under review). In the same task described above, I hypothesized that a speaker who is socially expected to speak precisely (a Nerdy person) will be interpreted as viewing an only approximate number on the phone screen less often than a person not expected to be precise (a Chill person); that is, I hypothesized the Nerdy persona condition to result in a higher proportion of covered selections than the Chill persona. Explicitly, the linking hypothesis was that a higher proportion of covered selections is the result of a higher rate of rejection of less precise matches when reasoning about the Nerdy persona, because a speaker assigned that persona is expected to be more precise.

All experimental materials, data, and analysis scripts for the experiment can be found in the author's github repository[1]; the experiment was also pre-registered, and this registration can be found at the Open Science Framework website[2].

[1] https://github.com/avelasqz/L245B_Project
[2] https://osf.io/54gyv

2 Experiment

2.1 Methods

2.1.1 Participants

170 participants were recruited via Prolific[3] for this experiment. Participant selection proceeded on the criteria that they be native English speakers, born in the United States, and currently residing in the United States. Participants were paid $4.04 to complete the task; average completion time was 14.5 minutes.

2.1.2 Materials

24 critical items were created, each with eight variations due to the 2 x 4 (persona: Nerdy or Chill; screen fit: match, near near match, near match, mismatch) design. The persona condition was between-subjects, so that each participant saw only one pair of speakers, both of which represented a single persona, either Nerdy or Chill (Figure 1). The screen fit condition was within-subjects, such that each participant saw 6 each of the match, near near match, near match, and mismatch conditions (Figure 2). The number on the visible screen in the match condition matched the utterance exactly in the distance and time conditions, and to the first decimal place in the cost condition; the near near match number deviated from the stated number by 5-11%; the near match by 12-18%; and the mismatch condition showed a number that deviated from the stated number by a large amount (by more than an hour and a half on time trials, and hundreds of miles or dollars in distance and cost trials). 8 of the critical items communicated cost; 8 communicated distance; and 8 communicated time.

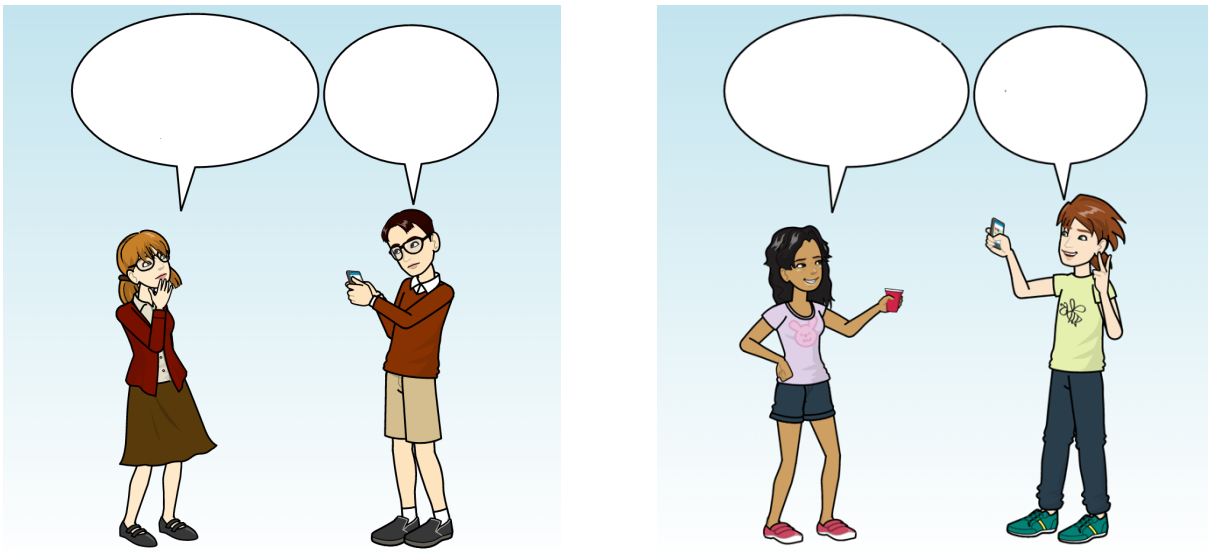

***Figure 1**: Rachel and Arthur (left) represent the Nerdy persona; Eva and Alex (right) represent the Chill persona.*

[3] app.prolific.co

***Figure 2**: From top left counterclockwise, the match, near near match, near match, and mismatch conditions for the response, "It's $300."*

24 filler items were also included, which showed two new characters engaged in a dialogue in which one character responded using the quantifier *some* (Figure 3). The uncovered phone screen in these cases showed a list of items that would make the *some* statement true, false, or underinformative; in this last condition, the underinformativity was consistent with the reading *some and possibly all*. An example of this last condition is shown in Figure 3.

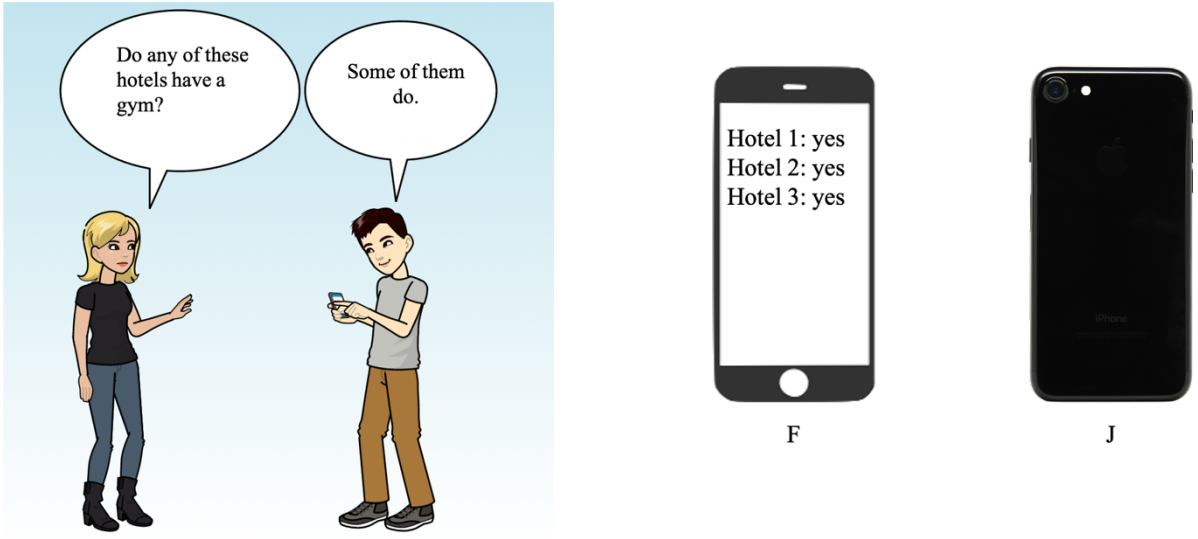

***Figure 3**: Becky and Tyler discuss hotels; the uncovered screen would be selected if the participant believed that Tyler's utterance was consistent with it (i.e., under a reading of 'some and possibly all').*

2.1.3 Procedure

Two manipulations were implemented in this experiment. The first was persona, a between-subjects factor, and the second was screen fit condition, a within-subjects factor. Each participant saw, then, only one pair of critical speakers (Rachel and Arthur or Eva and Alex) across 24 items, and the same 24 filler items featuring Becky and Tyler.

Participants first received the instructions in Figure 4, then completed an introductory trial, where the logic of the experimental task was laid out, and afterward three training trials, where participants were given feedback about their choice. After this, the critical and filler trials began, interleaved so that participants never saw two critical trials one after the other, nor two filler trials one after the other.

> We're going to play a little guessing game: you'll see some cartoon-like dialogues where one person asks another one a question.
>
> That person checks their phone and responds based on information they see there.
>
> You'll then be presented with two phone pictures. In one of them, you can see the details of the screen, while in the other one the phone is turned upside down.
>
> Your goal will be to guess which of the two phones is the one that the person is looking at, based on what this person is saying in the picture. As a general rule, you'll select the phone with the visible screen if you think that the information on the screen fits what is being said. You'll select the one turned upside down, instead, if you think that the information on the visible screen does **not** fit with what the person said.
>
> Note that **just one of the pictures** goes with what the second person responds. So you should only choose the phone that's turned upside down if you don't think the person giving the answer would have said what they did if they had been looking at the visible phone screen.

***Figure 4**: Experimental instructions.*

For each critical and filler trial, the scenario was first introduced in the top left corner, followed by a four-second pause, then an image of the characters with empty speech bubbles. After a three-second pause, the speech bubbles were filled, and after a further three-second pause, the question to be answered and the image of the phones was displayed on the right side of the screen. Participants registered their choice of phone by pressing the corresponding key (F or J). A complete display of what a participant saw on a single trial is shown in Figure 5.

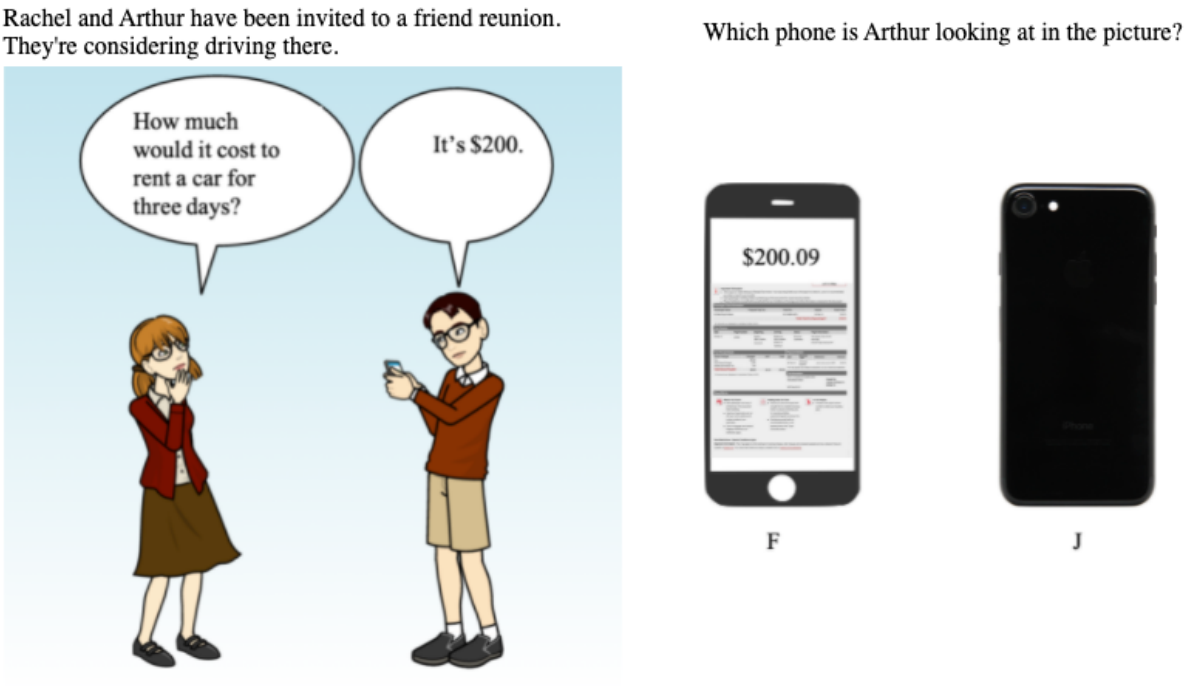

Figure 5: *A critical trial with Rachel and Arthur and in the match condition.*

In the original study, after the critical and filler trials were completed, two more questions were asked of participants: how nerdy they considered themselves to be, and how chill they considered themselves to be. However, this information was not collected in this replication.

2.2 Results

2.2.1 Variation in imprecision thresholds

Figure 6 shows the proportion of covered choices for each of the four conditions. This chart shows a difference between the match, mismatch, and two imprecise conditions (near near match and near match); in the mismatch condition, there appears to be a ceiling effect, so that the covered screen was almost always chosen when the displayed number differed significantly from the number uttered by the character, while in the match condition, we see covered responses at floor. The two imprecise conditions show rates between these two.

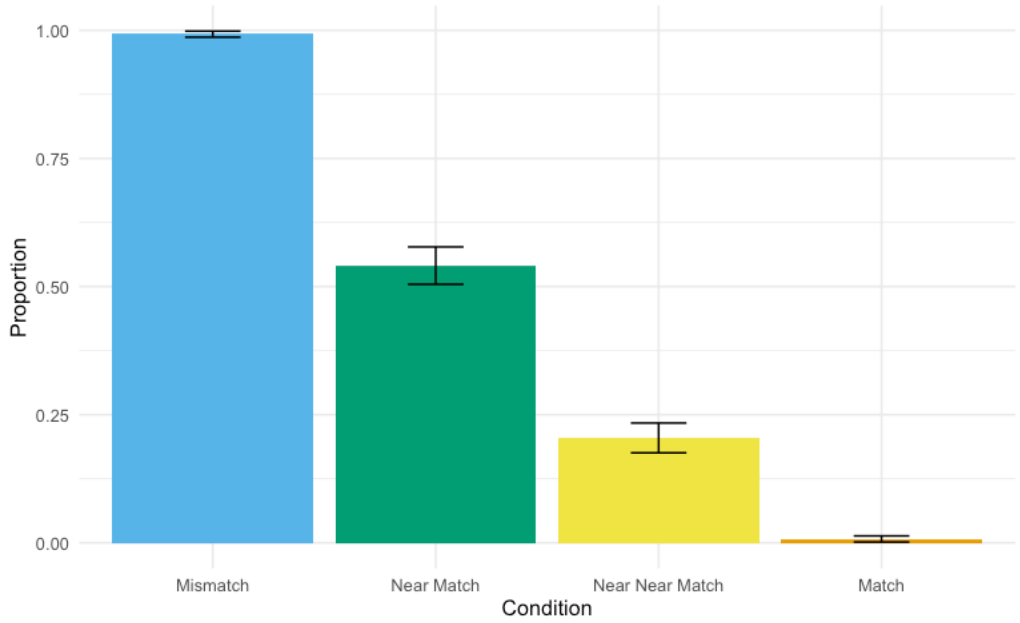

***Figure 6**: Proportion of covered responses by condition.*

I ran a mixed-effects logistic regression model (using the *glmer* function from the *lmerTest* package, as with all models herein; Kuznetsova et al. 2016) in R, with screen fit condition as a fixed effect and by-subject and by-item random intercepts with the imprecise condition as the reference level; models with random slopes did not converge. Table 1 gives the output for this model, and shows that the match and mismatch conditions differ significantly from the imprecise conditions. This confirms that there is a significant difference in responses between the imprecise conditions and the mismatch and match conditions.

Level	**Coefficient**	**Std. Error**	***p***
Intercept	-0.7513	0.1877	<.001
Match	-5.4163	0.4986	<.001
Mismatch	6.7342	0.4998	<.001

***Table 1**: Summary of mixed-effects logistic regression for imprecise, match, and mismatch conditions.*

2.2.2 Effect of persona on expectations of precision

Once the difference between the imprecise conditions, on the one hand, and the match and mismatch conditions, on the other, was established, I focused specifically on the imprecise conditions, near near match and near match, to determine the effect of persona on precision expectations. Figure 7 shows the proportion of covered choices for the near near match and near match conditions, by persona. The graph suggests that, in the near match condition, there were higher rates of covered choices for the Chill persona, and in the near near match condition, no difference in responses by persona. This is surprising; based on the original study, and on our own hypotheses, we should have seen higher proportions of covered choices with the Nerdy persona, in both near near match and near match conditions.

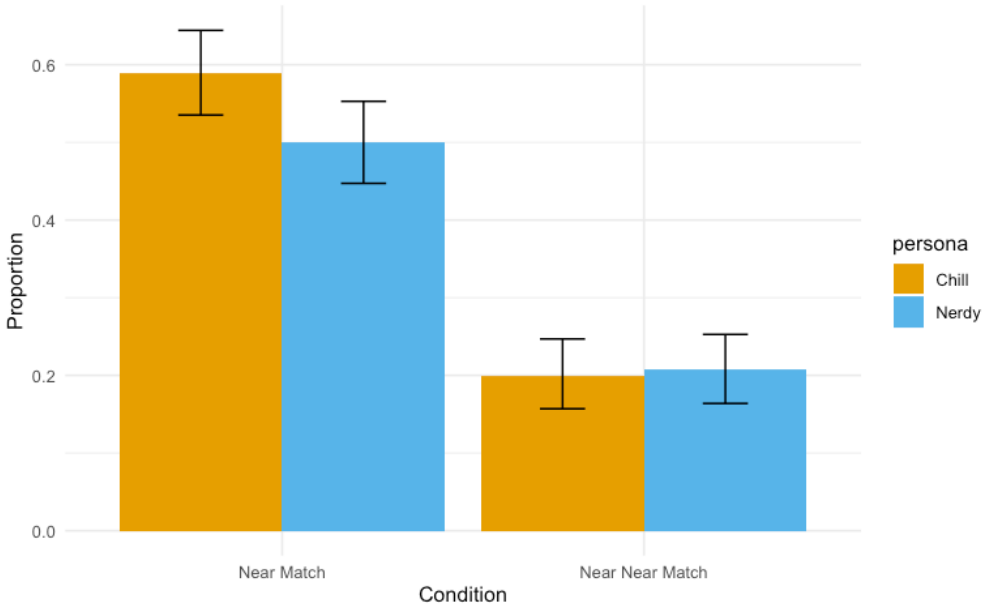

***Figure 7**: Proportion of covered responses by imprecise conditions and by persona.*

To explore these results statistically, I ran a mixed-effects logistic regression model with persona and condition, and their interaction, as fixed effects, by-subject and by-item random intercepts, and by-subject random slopes for the fixed effect of screen fit condition. Both condition and persona were centered. The results of this analysis can be found in Table 2.

Level	Coefficient	Std. Error	*p*
Intercept	-1.1925	.3070	<.001
Persona	-0.1722	.5192	.74
Screen Fit	-3.0390	.2993	<.001
Persona * Screen Fit	1.1047	.4595	<.05

***Table 2**: Summary of mixed-effects logistic regression for near match and near near match conditions.*

This model shows an overall bias of the data against choices of covered screen, as well as a significant effect of screen fit condition and of the interaction between screen fit and persona. What we do not find is a significant main effect of persona.

3 Discussion

The finding that screen fit condition had a significant effect on responses was to be expected, in that a number which differs from that uttered by the speaker by 12-18% should be accepted as the number referred to less often than a number which differs by only 5-11%. The interaction between persona and screen fit condition could be accounted for by a difference in the effect of condition for each persona, or in the effect of persona for each condition.

The lack of an effect of persona suggests two readily identifiable possible explanations. The first is that the effect of persona found by Beltrama and Schwarz was not replicated here, and may not be as reliable as that study suggested. However, a second explanation lies in the data collected by Beltrama and Schwarz about participant self-identification with relation to being Nerdy and Chill, and the effect this had on their results.

Beltrama and Schwarz asked participants to rate themselves on scales of Chill and Nerdy after the critical trials, and analyzed these in relation to the persona of the speakers encountered as ingroupness vs. outgroupness. In the experiment like the one replicated here, in which one group was recruited through Prolific and the other on the university campus, the online respondents showed a significant effect of the interaction between persona and ingroupness, such that a Nerdy participant had a lower proportion of covered screen responses than a Chill participant when evaluating a Nerdy speaker, and a Chill participant had a higher proportion of covered choices than a Nerdy participant when evaluating a Chill speaker. That is, the stereotypical expectations of Nerdy and Chill behavior seem attenuated—or even disappear—when an ingroup member is doing the evaluating.

Beltrama and Schwarz offer two possible explanations for this. The first is that all participants were able to access the stereotypical information in the same way, but that ingroup participants are less willing to make use of this information and describe a speaker like them with the use of stereotypes. The second is that stereotypical characteristics are more accessible to outgroup members, and therefore play more heavily in their reasoning about speakers they are unfamiliar with.

A similar phenomenon was found by D'Onofrio (2015), who looked at TRAP-backing and listeners' processing of ambiguous tokens in different persona and geographical region conditions. She found that speakers described as being Valley Girls or from California were expected to exhibit TRAP-backing at a higher rate, but that listeners who were not raised in the West were more likely to categorize ambiguous tokens as TRAP than those who were raised in the West. This raises the same questions as Beltrama and Schwarz' findings, namely, whether stereotypical information is accessed differentially depending on familiarity/solidarity with a group, or if the information is only used differently depending on group affiliation.

For the current results, what these findings mean is that the lack of effect of persona may be due to the distribution of participants who would identify as Nerdy or Chill if asked. That is, an unbalanced proportion of participants who would identify as Nerdy may have seen the Nerdy condition, and a large proportion of participants who identify as Chill may have seen the Chill condition. This could possibly cancel out the persona effect; unfortunately, we cannot investigate this further using the current data.

The impact of listener identity with relation to the speaker highlights an important facet of linguistic behavior which has sometimes been abstracted away from in theorizing and modelling: language behavior and knowledge is as much about the things being discussed as it is about the people doing the discussing. Sociolinguistics has for decades been addressing how social differences lead to meaningful variation in speech, and this has led most recently to the theorizing of social meaning and the personae constructed by speakers while speaking. Work in the third-wave variationist tradition has focused on communities of practice as the meaningful social unit for the social work done through language (see, e.g., Eckert 2000).

However, much semantic and pragmatic theorizing is still done on the level of the 'language', a construct that becomes suspect as a locus of investigation when differences such as those discussed above are considered. If there is differential access to stereotypical information and its effect on language processing, then there are real differences based on group

membership in the individual language system; if access is not differential, but only behavior, there is still a need to theorize and explain how this works in practice. And the challenge goes beyond group membership: given that individuals are members of a variety of groups simultaneously, and that very few individuals find themselves in the exact same place in the social landscape as any other, we see how differentiation in language systems and language behavior is found at the level of the individual. Yet, our theories and models often assume a monolithic 'English' or 'Spanish' or whatever other language across large populations of speakers, not taking into account differences among individuals and instead abstracting away from these.

Instead, linguistics may choose to tackle this challenge head-on, investigating language as the much more flexible system it is shown to be, and reap the benefits of a deeper and more complex understanding of language as human behavior and social practice.

Acknowledgements

My sincere thanks to Andrea Beltrama and Florian Schwarz for their openness about their project and for providing the experimental materials.

References

Acton, Eric, and Christopher Potts. 2014. That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics* 18(1): 3-31.

Beltrama, Andrea, and Florian Schwarz. Under review. Imprecision and speaker identity: How persona-level information affects meaning resolution.

Beltrama, Andrea, and Laura Staum Casasanto. 2017. totally tall sounds totally younger: Intensifiers at the socio-semantic interface. *Journal of Sociolinguistics* 21(2): 154-182.

D'Onofrio, Annette. 2015. Persona-based information shapes sociolinguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19(2): 241-256.

----- 2018. Personae and phonetic detail in sociolinguistic signs. *Language in Society* 47(4): 513-539.

Eckert, Penelope. 2000. *Language Variation as Social Practice*. Oxford: Blackwell.

Kuznetsova Alexandra, Per Bruun Brockhoff, Rune Haubo Bojesen Christensen (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26

Weatherholtz, Kodi, Kathryn Campbell-Kibler, and T Florian Jaeger. 2014. Socially mediated syntactic alignment. *Language Variation and Change* 26: 387-420.