Speaker-specific adaptation to variable use of uncertainty expressions

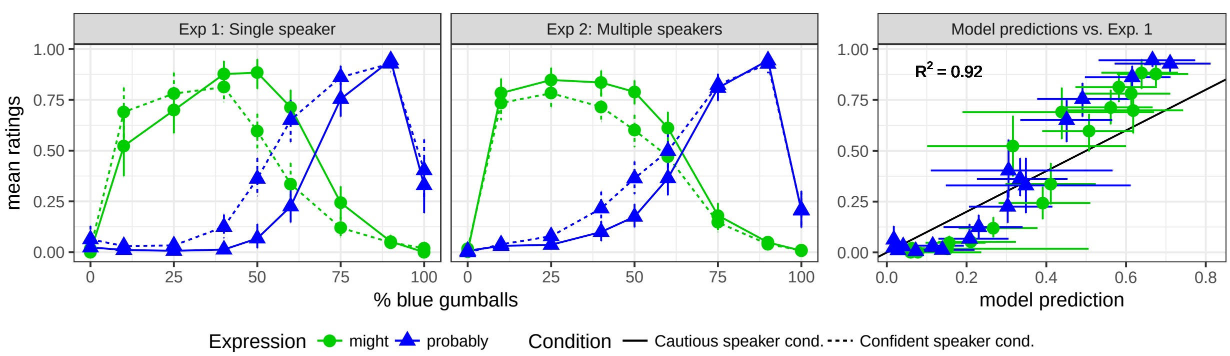
Sebastian Schuster (sebschu@stanford.edu) and Judith Degen, Stanford University

Speakers exhibit considerable production variability at all levels of linguistic representation [1-4]. This includes variability in lexical choice. For example, in the domain of quantifiers, when describing a bowl of candies in which half of the candies are blue and half are green, some speakers prefer the description "*Some* of the candies are green" over "*Many* of the candies are green" whereas other participants exhibit the opposite pattern [4]. Such speaker variability can be challenging for a listener who tries to infer what the world is like that the speaker is describing. Recent work across several linguistic domains has shown that listeners cope with this kind of variability by adapting to it and updating expectations [1-4]. We extend the research on semantic/pragmatic adaptation and investigate whether/how listeners adapt to varying uses of the uncertainty expressions *might* and *probably* (***"You might / You'll probably** get a blue gumball"*). In a norming study we found that speakers vary in their mapping between these expressions and the probability of getting a blue gumball. *Confident* speakers use *probably* if the objective probability of getting a blue gumball is 60%, whereas *cautious* speakers use *might.* How listeners should adapt to such variability is still poorly understood. We propose a novel **computational model** within the Rational Speech Act (RSA) framework [5] that treats adaptation as Bayesian update of listeners' beliefs about the speaker's underlying semantic thresholds, and we compare the model predictions to data from two experiments: listeners adapting to a single speaker (Exp. 1) and to multiple speakers (Exp. 2).

Exp. 1 (single-speaker adaptation): 61 MTurk participants saw 20 exposure trials (10 critical, 10 filler) followed by 36 test trials. Exposure trials showed a video of a speaker describing a gumball machine filled with blue and orange gumballs (critical trials: 60% blue gumballs). Participants in the *confident* and *cautious* condition were exposed to the speaker producing the *probably* and *might* form, respectively. Fillers were intended to boost trust in the speaker: on 5 trials, the speaker described a typical (25% and 90%, as estimated in the norming experiment) probability with the respective other uncertainty expression. On the other fillers, the speaker said of a 100% blue machine *"You'll get a blue one"*. On test trials, participants were asked to rate how likely they thought it was that the adult (who was shown in a still image) would produce the *might* or *probably* utterances or a blanket *something else* option. Participants distributed 100 points across these three options, for 9 blue gumball proportions. As the left panel below shows, *probably* was rated higher than *might* for a larger range of probabilities in the *confident* condition than in the *cautious* condition. Like [4], we quantified this difference by fitting a spline for each expression and participant and then computing the area under the curve (AUC). We found that the average difference between the AUC of *might* and *probably* is smaller in the *cautious* condition ($t(59) = -4.98$, $p < 0.001$), suggesting adaptation of expectations about the use of *might* and *probably.*

Exp. 2 (multiple speaker adaptation) investigated whether listeners formed *speaker-specific* expectations when they were exposed to two speakers whose uses of uncertainty expressions differ. 70 MTurk participants saw two exposure blocks, which both had the same structure as in Exp 1. In one of the blocks, participants saw a *cautious* speaker and in the other block, participants saw a *confident* speaker. The two exposure blocks were followed by two test blocks which probed listeners' expectations about the two exposure speakers' productions. The structure of each of these blocks was the same as in Exp. 1. In one of the blocks participants saw a picture of the *confident* speaker and in the other block they saw a picture of the *cautious* speaker when rating the three response options. As the center panel below shows, listeners formed speaker-specific expectations. We found that the average difference between the AUC of *might* and *probably* is smaller in participants' ratings in the *cautious speaker* block ($t(142) = -2.92, p < 0.01$), suggesting *speaker-specific* adaptation of expectations about the use of *might* and *probably*. However, as visible in the plots below, the adaptation effect was considerably smaller in this experiment (Cohens d: 0.486) than in the single speaker version (Exp. 1, Cohen's d: 1.263). Post-hoc analyses suggest that listeners were adapting to the speaker and the statistics of the situation independent of the speaker and that some listeners also adapted to the rating task leading them to provide consistent ratings across blocks.

Computational model: We model production as a pragmatic speaker S_1 within the RSA framework and assume the speaker reasons about a literal listener L_0 for which an utterance with an uncertainty expression u is semantically felicitous if the probability of getting a blue gumball exceeds a threshold θ_u [6-7]. We further assume that speakers have different preferences for different utterances modeled as costs c_u. Hence, the speaker's utterance choice depends on the thresholds θ and costs c. We estimate priors over these parameters from norming data and model adaptation as Bayesian belief updating of threshold and cost distributions (similarly as [1] and [8]). We predict participants' post-adaptation responses by exposing the model to the same observations as participants during the exposure phase. The right panel below shows that the model is able to predict participants' behavior accurately for the single-speaker experiment. For Exp. 2, the model makes the same predictions as for Exp 1. and hence the predictions are qualitatively correct but do not capture the smaller adaptation effect, which we observed independent of the task adaptation effects. A hierarchical model that models listener expectations as being guided by the statistics of specific speakers as well as the overall statistics of the experimental situation could predict the difference in adaptation effect sizes. Further, such a model would also make predictions about generalizations to novel speakers and situations and is the focus of our ongoing work.



References: [1] Kleinschmidt, D. & Jaeger, T.F. (2016). *Psychological Review.* [2] Kamide, Y. (2012). *Cognition.* [3] Fine, A.B., Jaeger, T.F., Farmer, T.A. & Qian, T. (2013). *PLOS ONE.* [4] Yildirim, I., Degen, J., Tanenhaus, M.K. & Jaeger, T.F. (2016). *JML.* [5] Goodman, N.D. & Frank, M.C. (2016). *TiCS.* [6] Lassiter, D. and Goodman, N.D. (2015). *Cognition.* [7] Herbstritt, M. and Franke, M. (2019). *Cognition.* [8] Qing, C. (2014). *MSc Thesis, UvA.*