‘Exhausting’ Theory of Mind resources impairs speaker-specific lexical alignment

Speakers can recognize inter-speaker variability in various pragmatic phenomena and adapt to the speakers’ different preferences of language (e.g.,[1],[2]). Furthermore, it has been repeatedly shown that interlocutors align with regard to their referential choices in what is commonly known as *lexical alignment* [3-6]. Moreover, we recently showed that in addition to alignment, individuals actively store speaker-specific lexical ‘stylistic’ choices, and that they use this knowledge to generalize speaker-specific information both in the linguistic and the social domains [7]. In this study, we aimed to examine the cognitive processes involved in the different stages of detecting, aligning with-, and generalizing speaker-specific language use. Specifically, we were interested in examining how these phenomena relate to (a) Theory of Mind (ToM), a social function and (b) Executive Functions (EF). It has been shown that performing cognitively demanding tasks can interfere with performance in subsequent language-related tasks [8]. Following this, we examined in this study whether performing a task that requires either using ToM (Reading the Mind in the Eyes Test (RMET)[9]), or inhibition-control (Flanker [10]) interferes with the ability to store, generalize, and align with speaker-specific language use.

Methods. Native Hebrew speakers (N=70, so far) took part in an online interactive picture selection task. Participants were led to believe they were engaging in an interactive task with other naïve participants. In fact, the ‘other participants’ were simulated by a computer program. Each participant was exposed to two different speakers, differing in their naming preferences for real-world objects, such that one speaker consistently produced disfavored words and the other one – their favored alternatives. Participants were assigned to one of three conditions. In one condition, participants performed the Reading the Mind in the Eyes Test before the experimental task; in the second condition, participants performed the Flanker task before the experimental task; the third condition was a control condition in which participants did not perform any task before the experimental task. In the experimental task, there were two different roles: *Directors* and *Matchers*. *Directors* instruct *Matchers* which image to choose, within an array of real-world objects. There were 5 steps in the task, always presented in the same order. (1) In the exposure phase, the participants acted as *Matchers* and were instructed by two other simulated participants (each in their turn) which image to choose. (2) In the alignment-test phase, participants acted as *Directors* and were required to instruct the simulated participants who were their directors (who supposedly now act as *Matchers*, each in their turn/block) which image to choose. (3) In the detection-test phase, participants were presented with an image on each trial and were asked if one of the simulated participants had used a certain word to describe the image. (4) The linguistic generalization phase included a task similar to the detection-test in which we asked participants if it is possible – hypothetically – that a given speaker would produce a certain utterance (of three different types - (a) common/uncommon adjective orders; (b) Sentences with non-canonical constituent order; (c) favored/disfavored words). (5) The social generalization phase included a rating task with a visual analog scale – for each speaker - asking about social and personality traits of each speaker (cooperation, book reading, number of friends, non-native language, and autistic traits).

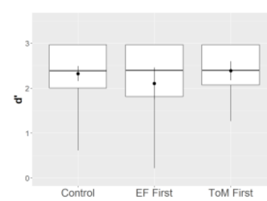

Figure 1. d' by condition.

Results. *Detection.* The ability to correctly map inter-speaker variability was analyzed using the d’ measure of the Signal Detection Theory [11], calculated per participant. In all conditions, the signal (speaker-word association) was reliably detected (control: $t(31) = 14.43$, $p < 0.001$; EF First: $t(19) = 8.03$, $p < 0.001$; ToM First: $t(17) = 18.3$; $p < 0.001$; Fig. 1) The d’ distributions did not significantly differ between the conditions.

Alignment. We fitted a mixed-effects logistic regression model predicting the odds of producing the less common alternative for each image by condition and speaker status. This model revealed a significant interaction (p < 0.001; Fig. 2), such that the odds of producing the disfavored word were higher when interacting with the uncommon speaker than with the common speaker, but only in the control (Z = 8.21; p < 0.001) and in the EF First (Z = 6.80; p < 0.001) conditions, and not in the ToM First condition (Z = 1.33; p = 0.18).

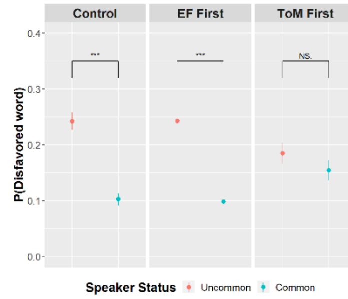

Figure 2. Probability of producing the disfavored word in the alignment-test phase by condition and speaker status.

Linguistic generalization. we conducted a separate analysis for each linguistic phenomenon, fitting three separate logistic regression models considering condition and speaker's status. We included only the uncommon forms of each phenomenon and analyzed the odds of accepting the association of each utterance to a given speaker (Fig. 3). For the lexical items, this model revealed that the odds of a positive response were higher for the uncommon speaker than for the common one, under all conditions. The other two phenomena did not reveal any significant effects.

Social generalization. We analyzed the ratings for each of the 5 questions separately. For each question, we fitted a mixed-effects ordered beta regression model predicting the numeric rating by condition and speaker status (Fig. 4). To sum up the results, in the control condition, we saw effects of speaker status and interactions for the cooperation, number of books, number of friends, and autism questions. These effects were absent in both the EF First and the ToM first conditions. To conclude, using ToM impairs speaker-specific lexical alignment, suggesting ToM is involved in this process. Furthermore, because social generalization was not observed in both the EF-First and in the ToM-First conditions, it seems that generalizing social information based on language-use requires available resources of both abilities.

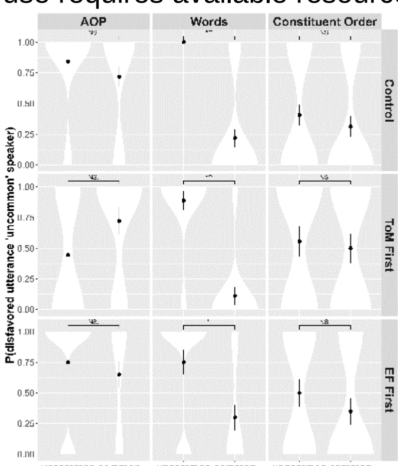

Figure 3. Probability of assigning the disfavored utterance to a speaker by condition and speaker status.

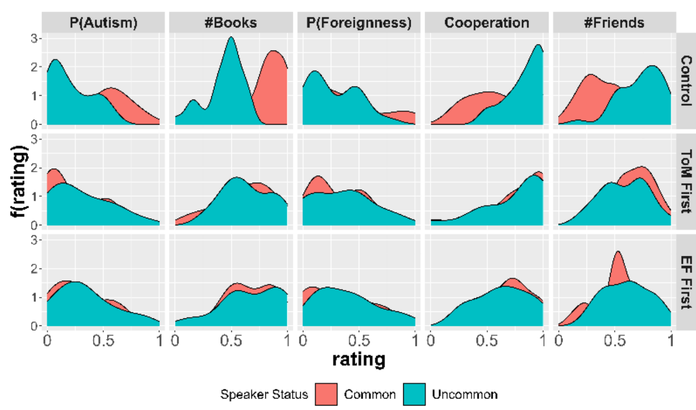

Figure 4. Ratings for each of the social questions, by condition and speaker status.

References. [1] Schuster, S., & Degen, J. (2020). *Cognition*. **[2]** Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). *Frontiers in psychology*. **[3]** Brennan, S. E., & Clark, H. H. (1996). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **[4]** Brown-Schmidt S. (2009). *Journal of memory and language*. **[5]** Clark, H. H., & Wilkes-Gibbs, D. (1986). *Cognition*. **[6]** Garrod, S., & Anderson, A. (1987). *Cognition*. **[7]** The authors. Under Review. **[8]** Saratsli, D., Trice, K. M., Papafragou, A., & Qi, Z. (2023). *Psyarxiv.* **[9]** Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. and Plumb, I. (2001). Journal of Child Psychology and Psychiatry **[10]** Eriksen, B.A., Eriksen, C.W. (1974). *Perception & Psychophysics*. **[11]** Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). *Psychological review*.