The role of reinforcement learning in pragmatic reasoning tasks:
Modeling individual differences in ACT-R

When playing as a comprehender in pragmatic reference games (RefGames), participants must determine the intended referent of a message using recursive Gricean reasoning about cooperative behavior (Frank & Goodman, 2012). Franke & Degen (2016) observe individual differences in RefGame performance: while some comprehenders do give responses consistent with second-order pragmatic reasoning (reasoning about a simulated cooperative speaker), others seem only to perform first-order reasoning (reasoning about a simulated literal speaker), or even avoid pragmatic interpretation altogether. Seeking to explain this variation, Mayn & Demberg (2023) correlated participants' performance on RefGame with a battery of individual difference tasks, and found a correlation with problem-solving tasks like Raven's Advanced Progressive Matrices (RAPM): participants who succeed in RAPM are more likely to give responses consistent with higher-order pragmatic strategies.

We suggest that this correlation arises because performance in RAPM and RefGame is dependent on shared domain-general resources: specifically, resources for reinforcement learning. Researchers have suggested that RAPM requires comprehenders to explore a wide space of possible rules and strategies (Carpenter et al. 1990; Vigneau et al. 2006; Gonthier & Roulin 2020). Stocco et al. (2021) use a cognitive model in the ACT-R framework to demonstrate that, in order to optimally explore this hypothesis space on each trial, a participant must deploy sufficient **negative feedback strength** (F_{neg}) (Frank et al. 2004) to quickly penalize unsuccessful strategies, so that they might identify a successful strategy before reaching the limits of their own personal **persistence** (Dale et al. 2018). The novel predictions of this model were borne out: a measure of F_{neg} based on the Probabilistic Stimulus Selection task (PSS, Frank et al. 2004) did predict higher RAPM performance.

An ACT-R model of RefGame

Inspired by the approach in Stocco et al. (2021), we present a model of RefGame using the ACT-R implementation `pyactr` (Brasoveanu & Dotačil, 2020) to demonstrate how performance may depend on these same reinforcement learning resources. The model selects referents by sampling with replacement from a space of three interpretation strategies, corresponding to interpretation which is (A) merely literal, (B) first-order pragmatic, or (C) second-order pragmatic. The model's simulated participants attempt these strategies on each trial, administering internal F_{neg} when a strategy fails to determine a unique optimal referent. They respond once a successful strategy has been found, or else they guess once persistence has been exhausted. Simulated participants thus only learn to apply second-order interpretation if self-administered feedback can accumulate over time enough to overcome a typical bias in favor of simpler response strategies.

As a preliminary validation, we assessed our model using the data from Mayn & Demberg (2023). Using `PyMC` (Abril-Pla et al. 2023, see Brasoveanu & Dotlacil, 2020), we fit F_{neg} parameters for the ACT-R models based on participants' performance, separately for RAPM and RefGame. These fitted F_{neg} values are positively correlated across the tasks ($R^2 = 0.10$, $p < 0.001$). If we model participants as using related F_{neg} across the tasks, we can indeed predict RefGame accuracy based on RAPM performance with some accuracy ($R^2 = 0.05$, $p = 0.02$).

Predictions of the ACT-R model

The model makes several other key predictions. (A) Because participants learn to engage pragmatic response strategies through experience with the task, selection of the intended targets should improve as the experiment progresses. (B) Because pragmatic response strategies involve extra reasoning about alternative messages, and alternative referents for alternative messages, they should take longer; thus, responses in critical trials which require pragmatic strategies will have longer RTs, and slower responses in critical trials will be more likely to be correct. (C) Individuals with higher F_{neg} and higher persistence will be more likely to deploy pragmatic response strategies, and thus will be more likely to respond accurately in critical trials (see Figure 1).

Experiment

We tested these predictions in a pre-registered[1] study on Prolific ($n = 150$). RefGame responses and response times were collected on 8 SIMPLE trials which only required first-order reasoning, and 8 COMPLEX trials which required second-order reasoning, following Franke & Degen (2016), along with 20 filler trials with unambiguous messages. Participants also completed individual differences tasks, including RAPM, the PSS measure of F_{neg}, and a 5-letter anagram task. Following Ventura & Shute (2013), we used this anagram task to assess persistence, calculated as a ratio of how long it took participants to decide to skip anagrams with no solution compared to their correct responses to easily solvable anagrams.

All effects discussed below were estimated in Bayesian mixed-effects models with regularizing priors in `brms` (Bürkner, 2017). As a quick estimate of the strength of the evidence for critical effects, we report the posterior probability P that $\beta > 0$ or $\beta < 0$ given the data.

Results

Results largely confirmed the predictions of the model. Re: (A), accuracy in critical items improved over trial order, for both SIMPLE ($P > 0.99$) and COMPLEX ($P = 0.95$), suggesting that success depends on learning optimal strategies over the course of the experiment. Re: (B), correct responses in both critical conditions were slower than on unambiguous fillers (Ps > 0.99); COMPLEX trials only trended as slower than SIMPLE ($P = 0.81$). We also observe that slower responses were more likely to be correct, most clearly in COMPLEX ($P = 0.95$). Finally, re: (C) we observe evidence that both higher F_{neg} and persistence, which showed no correlation with each other in our sample, were indeed associated with higher accuracy in SIMPLE ($P = 0.92$, $P > 0.99$) and COMPLEX ($Ps = 0.86, 0.91$) (see Figure 2). These did not account for the entirety of the (replicated) correlation between Raven's and RefGame, suggesting that there could be other relevant shared resources.

[1] https://osf.io/56gnf/?view_only=4b00464e8e4e4dc6b65d09243d219977

Discussion

Our results suggest that RefGame performance, and possibly pragmatic comprehension behavior more broadly, requires domain-general cognitive resources that help humans decide between multiple ways of solving a task. Individual differences in these resources may help explain some of the observed variation in pragmatic behavior in such tasks. We suggest that further development of cognitive models for pragmatic tasks is a fruitful way to make real-time accounts of pragmatic processing that are explicit and cognitively realistic.

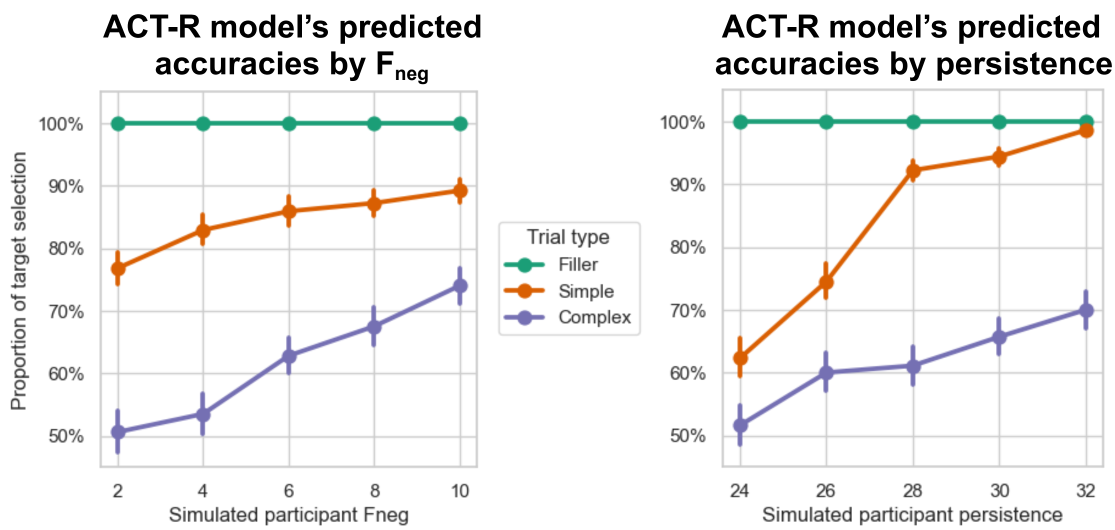

Figure 1: ACT-R model predictions for the relationship between RefGame accuracy and parameters corresponding to F_{neg} and persistence. Error bars correspond to bootstrapped 95% CIs.

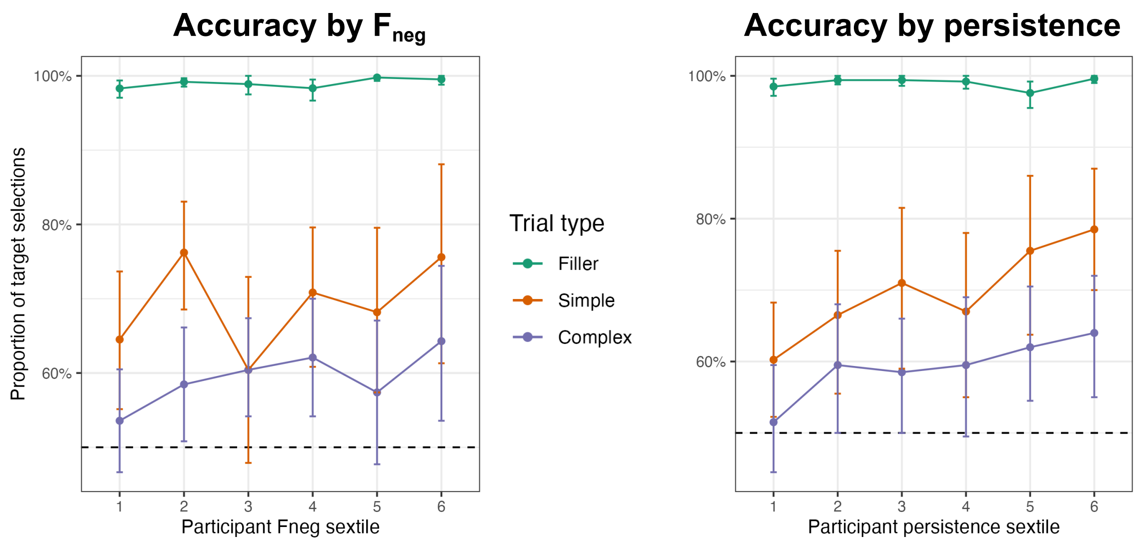

Figure 2: Observed relationships between RefGame accuracy and our measures of F_{neg} and persistence. Error bars correspond to bootstrapped 95% CIs.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science* 9: e1516

Brasoveanu, A., & Dotlačil, J. (2020). *Computational cognitive modeling and linguistic theory*. Springer.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. J*ournal of Statistical Software*, 80(1), 1-28.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404-431.

Dale, G., Sampers, D., Loo, S., & Green, C. S. (2018). Individual differences in exploration and persistence: Grit and beliefs about ability and reward. *PLoS ONE*, 13(9), e0203131.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, 306, 1940-1943.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998.

Franke, M., & Degen, J. (2016). Reasoning in reference games: Individual- vs. population-level probabilistic modeling. *PLoS ONE*, 11(5), e0154854.

Gonthier, C., & Roulin, J.-L. (2020). Intraindividual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition. *Journal of Experimental Psychology: General*, 149(3), 565-579.

Mayn, A. & Demberg, V. (2023). Sources of variability in a pragmatic reference game: Effects of reasoning, memory and perspective-taking. Poster presented at XPrag 10, Paris.

Stocco, A., Prat, C. S., & Graham, L. K. (2021). Individual differences in reward-based learning predict fluid reasoning abilities. *Cognitive Science*, 45, e12941.

Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, 29(6), 2568–2572.

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, 34, 261-272.