# Learning Pressures and Inductive Biases in Emergent Communication: Parallels between Humans and Deep Neural Networks

Lukas Galke[*1], Yoav Ram[2], and Limor Raviv[1,3]

[*]Corresponding Author: lukas.galke@mpi.nl
[1]LEADS group, Max Planck Institute for Psycholinguistics, Nijmegen, NL
[2]School of Zoology and Sagol School of Neuroscience, Tel Aviv University, Israel
[3]cSCAN, University of Glasgow, UK

Deep neural networks and humans are two types of learning systems with substantial differences in learning pressures. As many theories of language evolution rely heavily on learning pressures (Kirby et al., 2015; Smith & Kirby, 2008), it is currently unknown whether the learning pressures of humans are sufficiently reflected in deep neural network models in order to allow for insights to carry over and to advance theory building (Dupoux, 2018; Baroni, 2022). In emergent communication simulations, a population of deep neural networks starts from scratch without prior language knowledge and no predefined vocabulary, and are made to develop a language to solve a communication game via reinforcement learning (Lazaridou & Baroni, 2020). While these simulations have great potential for advancing our understanding of the emergence of languages, we can only expect insights gained with deep neural networks to inform language evolution research if the resulting AI languages show similar properties as natural languages (Brandizzi, 2023; Galke et al., 2022). Thus, finding and/or facilitating commonalities (i.e., by introducing appropriate inductive biases) can contribute to our understanding of how languages have evolved.

Reviewing the literature (Galke & Raviv, 2024), we find that the field of emergent communication has successfully designed models to replicate properties of natural languages, even when some of these properties were initially absent in such models. For instance, the lack of a least-effort bias in communicating neural network agents (Chaabouni et al., 2019, 2019; Lian et al., 2023), which gives rise to Zipf's law of abbreviation in natural languages (Kanwal et al., 2017; Zipf, 1949), can be addressed by inducing biases for lazy speakers and impatient listeners (Rita et al., 2020). When going to populations of communicating agents, another case is the absence of a group size effect (Chaabouni et al., 2022), i.e., that larger groups tend to develop more structured languages (Raviv et al., 2019), which can be (to some extent) addressed by introducing variation in learning rates (Rita et al., 2022) or by having agents alternate between sender and

receiver roles while restricting parameter updates (Michel et al., 2023). Most importantly, we find that a pressure for learnability, e.g., by having agents continually re-learning the language over and over again – modeled by resetting their parameters (Li & Bowling, 2019; Zhou et al., 2022) – seems to be indispensable for compositional structure to emerge consistently. This pressure for learnability closely resembles the iterated learning paradigm of language evolution research (Smith et al., 2003; Kirby et al., 2014). The necessity of re-learning for structure to emerge is commonly attributed to a learnability advantage of more compositional protocols – or conversely, the ease-of-teaching of compositional protocols to new agents (Li & Bowling, 2019). Although it has been shown for humans (Raviv, de Heer Kloots, & Meyer, 2021), this supposed learnability advantage of compositional structure for learning has not been tested with deep neural networks in a purely supervised learning setting.

Here, we test deep neural networks on their ability to learn new mini-languages with varying degree of compositional structure (Galke, Ram, & Raviv, 2023), analyzing whether more structure leads to more systematic generalization behaviour. We consider long short-term memory models (LSTM) (Hochreiter & Schmidhuber, 1997) trained from scratch as well as a large language model pre-trained on natural language (Brown et al., 2020; Ouyang et al., 2022). We ensure 1:1 comparability to humans by employing the same stimuli and procedure as in a previous study (Raviv et al., 2021). Our results show that – while all languages can be ultimately learned – more systematically structured languages, as quantified by topographic similarity (Brighton & Kirby, 2006), are learned better. Learning more structured languages also leads to more systematic generalizations to new, unseen items, and these generalizations are significantly more consistent and more human-like. Although differences in inductive biases between Transformers and LSTMs need to be taken into account (White & Cotterell, 2021), our findings lead to the clear prediction that children would also benefit from more systematic structure for learning – despite substantial differences in learning patterns compared to adults (Newport, 2020; Hudson Kam & Newport, 2005). This hypothesis is currently being tested (see preregistration: Lammertink et al. (2022)).

In summary, we have shown that deep neural networks display a learning and generalization advantage for more structured and compositional linguistic input – just as (adult) humans. This commonality between humans and machines, combined with other language properties facilitated by inductive biases in emergent communication, provides a rich testbed for using neural networks to simulate the very emergence of language in our species. In ongoing work, we seek to shed new light on why larger populations may tend to develop more structured languages. Notably, this group size effect has been shown to occur in humans even without iterated learning (Raviv et al., 2019), and we hypothesize that modeling cognitive constraints (e.g., memory) would bring us closer towards deep neural networks being useful models for studying human language evolution.

## References


Baroni, M. (2022). On the proper role of linguistically oriented deep net analysis in linguistic theorising. In *Algebraic structures in natural language.* CRC Press.

Brandizzi, N. (2023). Toward more human-like ai communication: A review of emergent communication research. *IEEE Access*, *11*, 142317-142340.

Brighton, H., & Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artif. Life*, *12*(2), 229–242.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020).*

Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2019). Anti-efficient encoding in emergent communication. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (pp. 6290–6300).

Chaabouni, R., Kharitonov, E., Lazaric, A., Dupoux, E., & Baroni, M. (2019). Word-order Biases in Deep-agent Emergent Communication. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5166–5175). Florence, Italy: Association for Computational Linguistics.

Chaabouni, R., Strub, F., Altché, F., Tarassov, E., Tallec, C., Davoodi, E., Mathewson, K. W., Tieleman, O., Lazaridou, A., & Piot, B. (2022). Emergent communication at scale. In *Iclr.*

Dupoux, E. (2018). Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59.

Galke, L., Ram, Y., & Raviv, L. (2022). Emergent communication for understanding human language evolution: What's missing? In *Emergent Communication Workshop at the Tenth International Conference on Learning Representations (EmeCom @ ICLR 2022).*

Galke, L., Ram, Y., & Raviv, L. (2023). What makes a language easy to deep-learn? *arXiv preprint abs/2302.12239*.

Galke, L., & Raviv, L. (2024). Emergent communication and learning pressures in language models: a language evolution perspective. *arXiv preprint abs/2403.14427*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Vari-

ation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151–195.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, *28*, 108–114.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Lammertink, I., Bazioni, M., Kloots, M. d. H., & Raviv, L. (2022, July). *Learnability effects in Children: are more structured languages easier to learn?* OSF.

Lazaridou, A., & Baroni, M. (2020). Emergent multi-agent communication in the deep learning era. *arXiv preprint abs/2006.02419*.

Li, F., & Bowling, M. (2019). Ease-of-teaching and language structure from emergent communication. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (pp. 15825–15835).

Lian, Y., Bisazza, A., & Verhoef, T. (2023). Communication Drives the Emergence of Language Universals in Neural Agents: Evidence from the Word-order/Case-marking Trade-off. *Transactions of the Association for Computational Linguistics*, *11*, 1033–1047.

Michel, P., Rita, M., Mathewson, K. W., Tieleman, O., & Lazaridou, A. (2023). Revisiting Populations in multi-agent Communication. In *The Eleventh International Conference on Learning Representations (ICLR 2023).*

Newport, E. L. (2020). Children and Adults as Language Learners: Rules, Variation, and Maturational Change. *Topics in Cognitive Science*, *12*(1), 153–169.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv preprint abs/2203.02155*.

Raviv, L., de Heer Kloots, M., & Meyer, A. (2021). What makes a language easy to learn? a preregistered study on how systematic structure and community size affect language learnability. *Cognition*, *210*, 104620.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Rita, M., Chaabouni, R., & Dupoux, E. (2020). "lazimpa": Lazy and impatient neural agents learn to communicate efficiently. In *Proceedings of the 24th Conference on Computational Natural Language Learning (CoNLL 2020)*

(pp. 335–343). Association for Computational Linguistics.

Rita, M., Strub, F., Grill, J.-B., Pietquin, O., & Dupoux, E. (2022). On the role of population heterogeneity in emergent communication. In *The Tenth International Conference on Learning Representations (ICLR 2022)*.

Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3591–3603.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial life*, *9*(4), 371–386.

White, J. C., & Cotterell, R. (2021). Examining the Inductive Bias of Neural Language Models with Artificial Languages. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 454–463). Online: Association for Computational Linguistics.

Zhou, H., Vani, A., Larochelle, H., & Courville, A. (2022). Fortuitous Forgetting in Connectionist Networks. In *The Tenth International Conference on Learning Representations (ICLR 2022)*.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*.