Précis of *The Cultural Emergence of Combinatorial Structure*

Matthias Hofer

Cultural symbol systems, such as language, music, or visual schemas and diagrams, are fundamental to human life, enabling us to communicate, collaborate, and solve problems in wide variety of domains. Language, for example, allows us to describe and coordinate the actions needed to build a shelter or find food. Diagrams and schemas serve as tools to visually represent complex, abstract ideas, facilitating collaboration and problem-solving. Music can evoke deep emotional responses in listeners and create a sense of shared identity. Finally, these systems enable the storage and transmission of knowledge across generations, a fundamental cornerstone of human culture (Deacon, 1998).

Central to the effectiveness of these systems is their *combinatorial structure*. From phonemes in language, to notes in a melody, or the lines in a diagram, these systems combine a finite set of building blocks to express a set of concepts or ideas. This flexible recombination of elements, as illustrated by the versatile use of phonemes in languages to form countless words (Yule, 2022), exemplifies the productivity that is characteristic of combinatorial systems (O'Donnell, 2015).

The goal of this thesis is to provide a more comprehensive account of the evolution of combinatorial symbol systems, following an interdisciplinary approach that weaves together insights from experimental psychology, psycholinguistics, evolutionary science, and computational modeling. Starting point for investigation is a puzzling observation: In human culture, combinatorial structures are predominantly found in language, including speech, gestures, and writing systems, as well as in music. This contrasts with an abundance of non-combinatorial signaling systems in human culture—such as diving signals, traffic signs, sports referee signals, and hazard symbols—where combinatoriality seems feasible but is absent. Why do we not see a more widespread use of combinatorial systems?

To address this, the thesis will argue that combinatorial structure only emerges under a set of very specific conditions. Accordingly, the research is structured into the following distinct sections:

- It is suggested that one challenge in the evolution of combinatorial systems is **overcoming iconicity**, where signs or gestures possess inherent, specific meanings that resist free recombination. Iconicity offers an alternative solution to problems related to learning and transmission that arise in cultural evolution. The first study tests the hypothesis that iconicity and combinatorial structure might conflict.
- The second section hypothesizes that combinatorial structure emerges from balancing simplicity and informativeness, placing emphasis on the the **role of communicative efficiency**, alongside the established role of learning bottlenecks in the emergence of structure. This offers insights into the specific conditions under which combinatoriality might arise in human culture.
- The final section addresses the **computational basis** for the human capacity for combinatorial structure. It elaborates on previous speculations about simplicity-focused inductive biases that arise in human learning by leveraging advancements in probabilistic programming and machine learning. A neuro-symbolic computational model is developed, combining opposing approaches to concept learning and generalization. This provides evidence for the role of symbolic inductive biases in the emergence of structured symbol systems.

Taken together, these individual studies help motivate a broader cohesive vision: Combinatorial systems emerge when specific cognitive processes, underpinned by computational adaptations for symbolic processing, are amplified by social-cultural processes on timescales that extend beyond the individual learner. This synthesis not only addresses a fundamental question in the evolution of human communication, but also exemplifies broader patterns in cultural development and cognitive science.

Overcoming Iconicity

A central question in evolutionary linguistics is the relationship between iconicity and combinatoriality. Iconicity, where symbols reflect sensory-motor attributes of their referents, reflects a departure of the principle of arbitrariness central to linguistics (Mahowald, Dautriche, Gibson, & Piantadosi, 2018; De Saussure et al., 1916). Crucially, iconicity seems to conflict with combinatorial structure. The latter requires that symbols be formed using a small set of building blocks, conforming to syntactic rules rather than semantic properties, enabling the system to generate limitless expressions from a finite set of elements (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Roberts, Lewandowski, & Galantucci, 2015). Iconicity, in contrast, imposes constraints on the relationship between form and meaning. This can be challenging in a combinatorial system where symbols are constructed from a predefined set of building blocks, potentially limiting the ability to directly mirror relevant referent features.

Research further indicates that iconicity and combinatoriality may fulfill distinct roles. Iconicity is foundational for bootstrapping communication, particularly in the absence of established conventions. Studies show that in the early stages of language development, gestures, which have a stronger affinity for iconicity, are preferred over spoken forms due to their representational transparency (Fay et al., 2022). Combinatoriality, on the other hand, seems to emerge for transmission efficiency, making messages themselves simpler to produce and perceive (Roberts et al., 2015). This suggests that signals initially start out iconic but will eventually become more arbitrary as they develop into combinatorial systems (Zuidema & de Boer, 2018). Consistent with this, Verhoef, Kirby, and de Boer (2016) conducted an experiment where participants transmitted artificial languages. Although the direct contribution of iconicity could not be measured, the authors found that the onset of combinatorial structure was delayed in a condition where forming iconic relationships was possible.

This chapter aims to bridge this research gap by providing quantitative evidence of the possible tension between iconicity and combinatorial structure. Additionally, it focuses on a more abstract, systematic form of iconicity that exists in terms of the perceptual complexity of meanings and signals (Lewis & Frank, 2016). This approach, which considers the compatibility of iconicity with combinatorial building blocks, presents an opportunity to develop a more nuanced account that considers a more mutual rather than a conflicting relationship.

To explore the simultaneous evolution of combinatoriality and iconicity, an iterated learning experiment (Smith, Tamariz, & Kirby, 2013) was designed using artificial languages comprised of whistled signals. This signaling space was chosen based on pioneering work by Verhoef, Kirby, and de Boer (2014). What makes this signaling modality particularly compelling is its similarity to fundamental aspects of the vocal-auditory modality in human speech, while simultaneously being a novel medium in which most people lack the same level of domain expertise as they have in language. In the experiment, participants formed 'transmission chains' and learned whistled signals of varying complexity (illustrated in Figure 1), which initially contained no combinatorial elements. These signals were arbitrarily paired with visual referents of different complexity in a non-systematic way. Each participant then transmitted the language to the next in line, who acquired it in the same way they did. The languages underwent evolution over 10 generations. The development of combinatoriality and iconicity was then evaluated on the basis of a set of separate experiments, which assessed complexity of the evolved signals and languages as a whole.

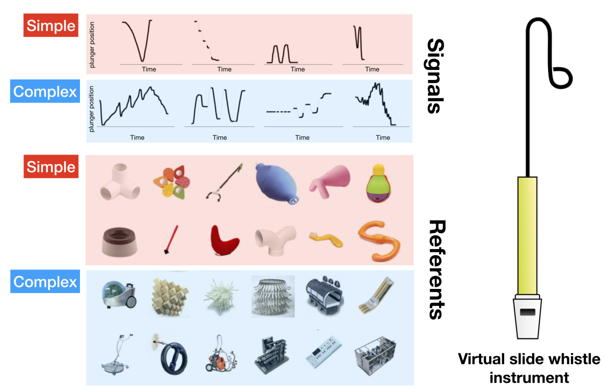

Figure 1: Illustration of the experimental stimuli and the virtual slide whistle instrument. The graphic shows the whistled signals, with lines indicating pitch trajectories, alongside visual referents, categorized as simple (red) or complex (blue). Initially, these were paired in an unsystematic manner so that languages would start out as non-iconic.

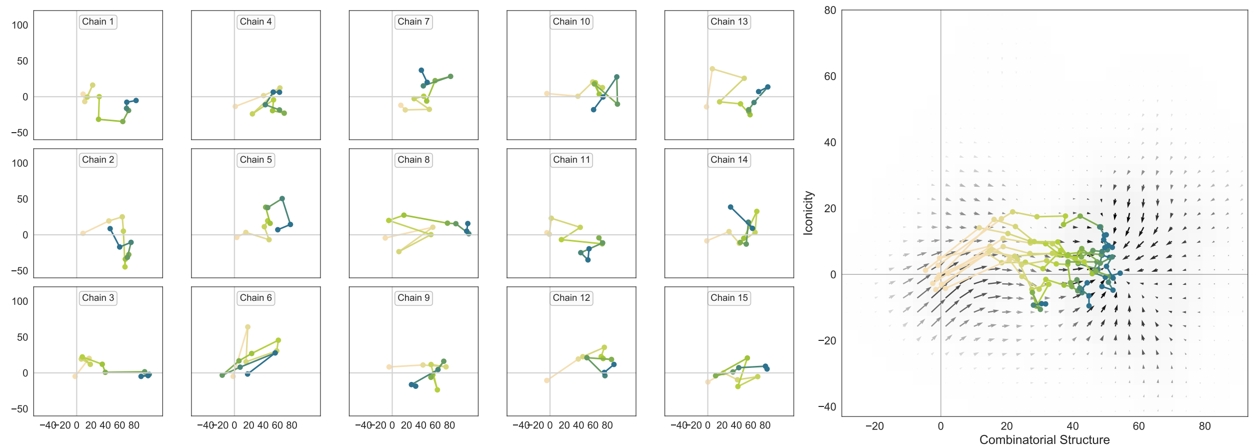

Figure 2: Visual representation of the evolutionary trajectories of all fifteen experimental chains, on the basis of which we constructed a vector field representation that shows inter-generational parameter changes for observed and hypothetical languages. TODO: add more detail on methods and add figure axis labels

Results indicated a gradual emergence of combinatorial structure across multiple generations. Notably, iconicity was rapidly incorporated but eventually diminished to levels indistinguishable from arbitrariness. This trend is depicted in Figure 2, presenting a vector-based representation of how these two properties evolved in relation to each other. This decline aligns with the conflict account, suggesting a state-dependent interplay between combinatorial structure and iconicity.

These results support the 'conflict account' outlined above, while also presenting a challenge that demands further research. As predicted under the account, iconicity diminishes as combinatorial structure continues to develop. This could be attributed to the favorable conditions for combinatorial structure in this setting, where subjects are required to encode, recall and accurately transmit a set of arbitrary and diverse signals, especially considering the absence of actual communicative use of these signals. On the other hand, the disappearance of this more abstract form of iconicity, characterized by complexity relations,is intriguing. Theoretically, such a form of iconicity could coexist with different combinatorial structures, making its disappearance unexpected and challenging the mere conflict account. The loss of iconicity, even in forms potentially compatible with combinatorial structures, suggests a complex interplay among the underlying cognitive biases and their expressed preferences for one linguistic feature over another at different stages.

Balancing Simplicity and Communicative Efficiency

This chapter of the thesis shifts focus from iconicity to the mechanisms behind the emergence of combinatorial structure. It aims to explain its emergence as a result of conflicting forces acting cultural evolution, an idea with historical roots in linguistics (Zipf, 1949; Martinet, 1952; Hockett, 1960) and recent grounding in mathematical accounts building on information theory (Plotkin & Nowak, 2000; Gibson et al., 2019).

Linguistic structure is thought to emerge from trade-offs between simplicity and informativeness, amplified across generations through cultural transmission. Learning pressures favor simpler languages (Smith, Kirby, & Brighton, 2003). However, overly simplistic languages risk losing communicative utility. To maintain informativeness, languages must balance simplicity with the need to make clear distinctions, effectively limiting excessive simplification (Spike, Stadler, Kirby, & Smith, 2017). Combinatorial structure seems to offer a solution to this balance, an idea that has recently been corroborated through a computational model presented in Kirby and Tamariz (2022). By reusing elements in systematic ways, languages can remain simple (easier to learn and use) yet expressive (capable of conveying diverse messages). This suggests combinatorial systems are preferred over holistic ones, as they align better with human learning biases, making them more efficient.

This chapter aims to empirically test this trade-off hypothesis. An online multiplayer experiment was developed which expands on the iterated learning framework introduced in the first chapter. This is done by incorporating a communicative task, allowing the examination of signaling systems under simultaneous pressures for simplicity and informativeness organically arising from the communicative task. The same whistled signal modality is used as in the previous set of studies.

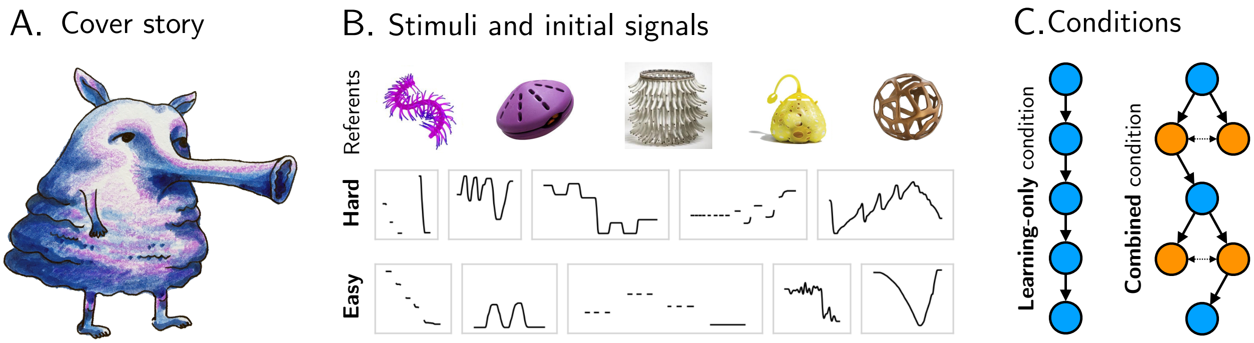

Figure 3: **A**. In the cover story, subjects were told that they learn signals produced by aliens on planet Vuma, who communicate using their whistle-like trunk. **B**. Data for two separate initial conditions was collected, where either five individually simpler or five slightly more complex signals were paired with novel visual referents taken from (Lewis & Frank, 2016). The assignment of referents to signals was randomized across chains and between successive generations. **C**. The two different experimental conditions.

The study contrasts a learning-only condition and a combined-condition where languages were repeatedly learned and used communicatively before they were transmitted to the next participant (see Figure 3). TODO: add more detail about stimuli and methods

It is found that, crucially, only languages evolving under both learning and communication develop system-wide combinatoriality (as illustrated in Figure 4). Communicative played a crucial role in allowing the reduction of complexity without compromising signal discriminability, leading to informative systems adapted for efficient communication. It also facilitated the alignment of individual combinatorial patterns into broader, shared conventions, enhancing communication performance.

We might expect the communicative task to also contribute toward signal simplification (speakers want to say as little as possible to reduce effort). This approach of characterizing languages in terms of efficiency concerns arising solely in communication has been used to understand the typological diversity of languages, without considering the role of learning biases or the process of language acquisition (Kemp & Regier, 2012; Zaslavsky, Kemp, Regier, & Tishby, 2018; Zaslavsky, Garvin, Kemp, Tishby, & Regier, 2021). Interestingly, contrary to this expectation, the communicative task did not simplify individual signals as anticipated but slightly increased overall language complexity. This suggests that while simplicity pressures during acquisition act on the system level, communicative pressures during usage can lead to complexity at the signal level.

These results support the hypothesis that domain-general cognitive biases for simplicity, coupled with functional pressures for informativeness, can shape the evolution of communication systems. The emergence of combinatorial structure in language appears to be a solution to the trade-off between simplicity and informativeness, rather than an inevitable feature of symbol systems. Combinatorial systems are simple due to element reuse, yet informative due to the diverse signals created by rearranging these elements. This helps us better understand the puzzle outlined in the introductory section. TODO: work on the final paragraph

The Computational Basis of Combinatorial Structure

Following the exploration of issues of iconicity and the role of communicative efficiency, the final thesis chapter delves into the cognitive underpinnings that enable such systems to evolve. A critical question that emerges is: What are the cognitive adaptations in human learners that facilitate the development of combinatorial systems?

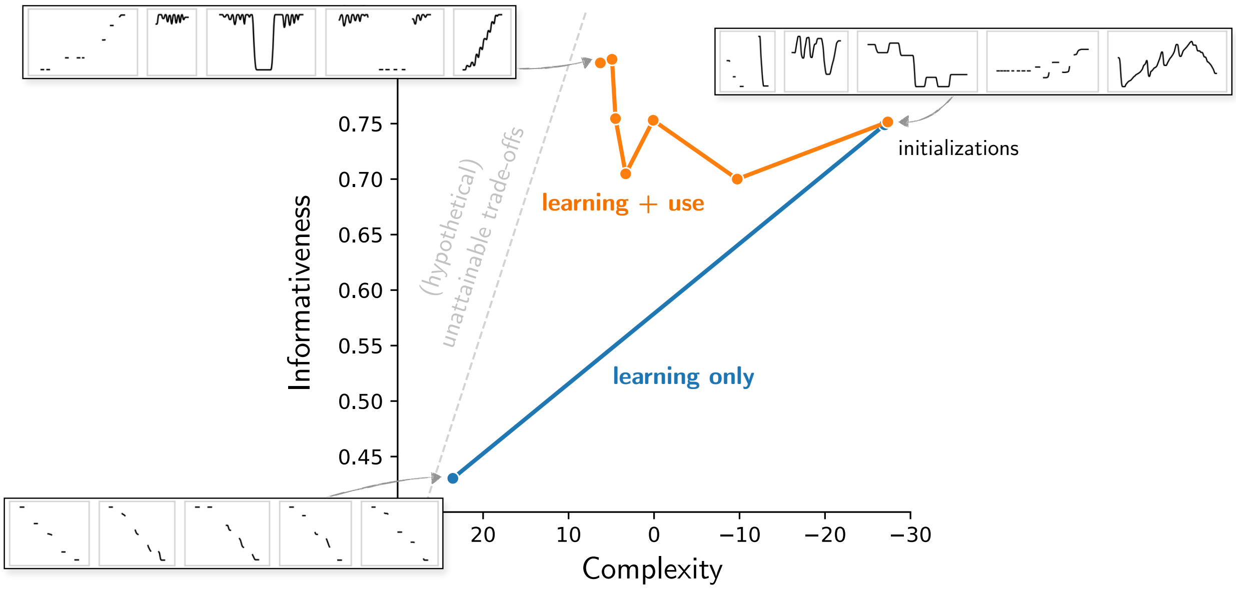

Figure 4: Results visualizing trade-offs between simplicity and informativeness. Languages start out as both complex and expressive (example language in the top right corner). Languages in the combined condition increase in communicative utility and decrease in complexity. This reduction in complexity is achieved through the utilization of combinatorial building blocks, as exemplified by the top left language in the figure. Languages in the learning-only condition become even simpler over time, but do so at the expense of loosing discriminability (see the example language in the bottom left corner).

The human propensity for detecting combinatorial patterns, even in the absence of explicit structure, is well-documented. This is particularly evident in the spontaneous emergence of combinatorial patterns in home sign systems by deaf children and the development of Al-Sayyid Bedouin Sign Language (ABSL) (Senghas & Kita, 2004; Aronoff, Meir, Padden, & Sandler, 2008). These cases highlight the importance of understanding the learning constraints and computational mechanisms during the acquisition of such systems.

Our approach to exploring these cognitive adaptations involves developing computational models that mimic human learning biases in combinatorial systems. For this we draw on work in the domain of concept learning, where various model classes ranging from structured probabilistic models (Lake, Salakhutdinov, & Tenenbaum, 2015) to neural networks (Ganin, Kulkarni, Babuschkin, Eslami, & Vinyals, 2018) have been proposed. These models, however, have limitations in fully capturing the emergence of new symbol systems or scaling to domains beyond their initial scope. We propose a neuro-symbolic generative model integrating the strengths of symbolic approaches with the flexibility of neural models. This model aims to provide a coherent computational account of how combinatorial systems are learned and, ultimately, how they evolve.

Our neuro-symbolic generative model treats signals as sequences of discrete segments, akin to linguistic phonemes, and uses neural networks to render these in a continuous signal-transmission modality. This approach allows for systematic generalization typical of symbolic models, while also handling continuous perceptual data. We compare our model to neural network-only architectures, such as recurrent neural networks (RNNs) and transformer models, focusing on their ability to learn and generalize from whistled signal data obtained from previous experiments discussed earlier.

Our findings demonstrate that the neuro-symbolic model outperforms neural network-only models in learning and generalizing the whistled signal data, and also behaves more human like in terms of the errors it makes. This suggests that the human ability to develop combinatorial systems may be rooted in a blend of symbolic reasoning and neural processing. The implications of this research extend beyond linguistics to broader cognitive science domains. Understanding the cognitive mechanisms that underlie the acquisition and evolution of knowledge domains such as language that seem to be characterized by rich and systematic

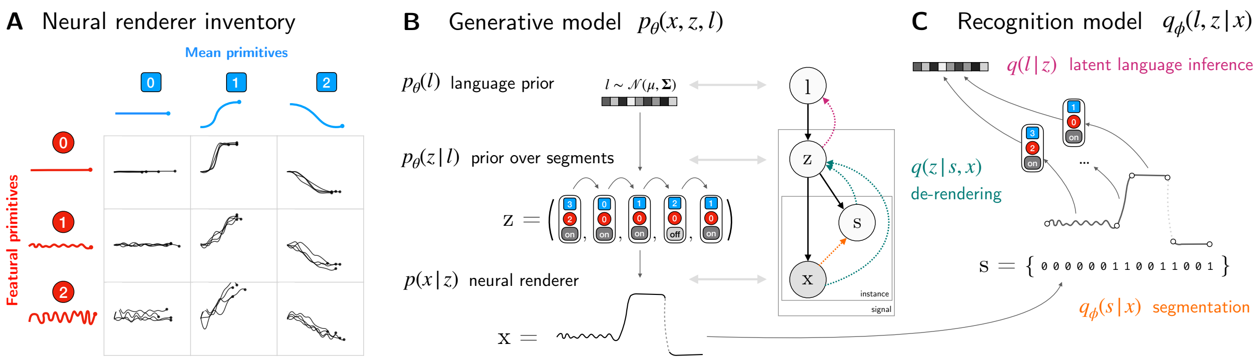

Figure 5: **Neuro-symbolic generative model.** (A.) A neural renderer, which consists of an inventory of mean and featural primitives that compose into rendered signal parts. (B.) Starting from these primitives, signals in the generative model are composed using a prior over segments $z \sim p_\theta(z)$ and a neural renderer $p(xz)$ which produces the continuous signal x. The hierarchical model additionally consists of a global latent code ℓ which parameterizes the bigram conditional prior $p_\theta(z\ell)$. (D.) The recognition model used for inference is decomposed into *segmentation* $q_\phi(sx)$ (either filtering- or neural network-based) and derendering $q(zs, x)$.

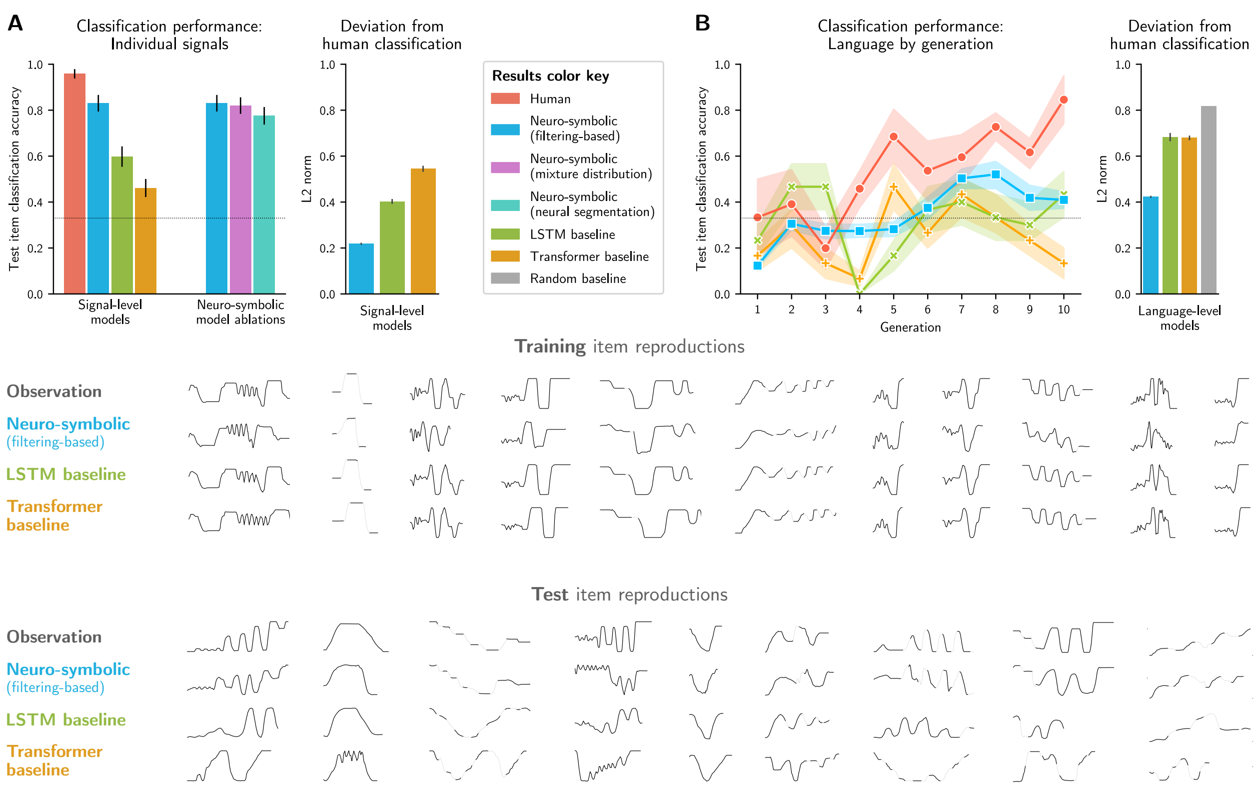

Figure 6: (A.) The neurosymbolic model classifies test set items more accurately and its performance is more human-like than either neural baseline. There is no significant difference in performance between the different inference models. (B.) In language classification, our hierarchical neurosymbolic model captures human performance better than the two neural baseline models. Like in the human subject experiment, neurosymbolic classification accuracy increases as a function of generation, indicating that the model successfully picks up combinatorial structures from more mature languages.

causal relationships but are grounded in perception is key to comprehending human intelligence. Our model provides insights into how structured concepts are acquired from raw data, a challenge in current machine learning systems.

Conclusion

One central questions in cognitive science is to understand the origins and development of combinatorial symbol systems in human culture, enabling diverse forms of communication, collaboration and creativity. Among important outstanding questions (Zuidema & de Boer, 2018), this thesis has specifically focused on the social and cultural processes that might allow these systems to evolve, as well as the cognitive capacities that enable us to acquire and use them effectively. By weaving together novel behavioral and computational evidence, the thesis offers insights into how combinatorial structure could have evolved. These results also shed light on the puzzling observation that combinatorial structure, despite its apparent utility in communication and representation, is so rare in natural systems:

- **Overcoming iconicity.** The initial study demonstrates a natural progression from iconicity to combinatorial systems, suggesting that the two might be in conflict. The ready adoption of iconicity, even though it was totally absent in the input, suggests that combinatorial structure might not be immediately prevalent or may not evolve at all in contexts where the benefits iconicity outweigh those of combinatoriality.
- **The role of communicative efficiency.** The second study provides key evidence that combinatorial systems likely emerge only under conditions where simplicity and informativeness (among other factors) are optimally balanced, as demonstrated through a cultural transmission experiment that incorporates a reference game.
- **The computational basis of combinatoriality.** Computational modeling in the final part of the thesis reveals the cognitive biases necessary for a learner to develop combinatorial structures. Beyond a mere preference for simplicity, it suggests that learners must possess the capacity to analyze and represent inputs symbolically. This requirement could explain the scarcity of combinatorial structures in other species lacking such sophisticated symbolic processing abilities.

The thesis revisits key themes such as the tension between iconicity and combinatorial structure, the role of meaning and compositionality in symbol systems, and the hierarchical nature of signaling systems. It discusses how simplicity preferences and iconicity manifest differently at individual signal and system levels, and how evolutionary pressures might influence these systems.

While the thesis makes significant contributions to our understanding of combinatorial symbol systems, several important limitations need to be acknowledged, such as the use of a simplified auditory signaling space and the potential influence of pre-existing knowledge of combinatorial patterns. Future research should explore the emergence of combinatorial systems in more diverse modalities (Little, Eryılmaz, & de Boer, 2017) and in more naturalistic and complex settings (Morin et al., 2018). Additionally, while focusing on human symbol systems, the thesis recognizes that combinatorial signaling is not unique to humans. Future studies should compare human capabilities with those of other species (Engesser & Townsend, 2019) and examine the influence of social and cultural factors, such as community size and structure (Kirby & Tamariz, 2022), on the evolution of combinatoriality.

Especially the neurosymbolic modeling framework presented in this thesis would be ideally suited to integrate perspectives on the biological evolution of the basis for combinatoriality and compare the combinatorial capacities of humans with those of other species. Understanding these aspects will further elucidate the unique position of human combinatorial structures in the broader tapestry of communication and cognition.

References

Aronoff, M., Meir, I., Padden, C. A., & Sandler, W. (2008, March). The roots of linguistic organization in a new language. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *9*(1), 133–153. doi: 10.1075/is.9.1.10aro

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain* (No. 202). WW Norton & Company.

De Saussure, F., et al. (1916). Nature of the linguistic sign. *Course in general linguistics*, *1*, 65–70.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015, October). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. doi: 10.1016/j.tics.2015.07.013

Engesser, S., & Townsend, S. W. (2019, February). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493. doi: 10.1002/wcs.1493

Fay, N., Walker, B., Ellison, T. M., Blundell, Z., De Kleine, N., Garde, M., ... Goldin-Meadow, S. (2022, March). Gesture is the primary modality for language creation. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1970), 20220066. doi: 10.1098/rspb.2022.0066

Ganin, Y., Kulkarni, T., Babuschkin, I., Eslami, S. A., & Vinyals, O. (2018). Synthesizing Programs for Images using Reinforced Adversarial Learning. In *International Conference on Machine Learning* (pp. 1666–1675).

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019, May). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. doi: 10.1016/j.tics.2019.02.003

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Kemp, C., & Regier, T. (2012, May). Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, *336*(6084), 1049–1054. doi: 10.1126/science.1218811

Kirby, S., & Tamariz, M. (2022, January). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200319. doi: 10.1098/rstb.2020.0319

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015, December). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. doi: 10.1126/science.aab3050

Lewis, M. L., & Frank, M. C. (2016, August). The length of words reflects their conceptual complexity. *Cognition*, *153*, 182–195. doi: 10.1016/j.cognition.2016.04.003

Little, H., Eryılmaz, K., & de Boer, B. (2017, November). Signal dimensionality and the emergence of combinatorial structure. *Cognition*, *168*, 1–15. doi: 10.1016/j.cognition.2017.06.011

Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018, November). Word Forms Are Structured for Efficient Use. *Cognitive Science*, *42*(8), 3116–3134. doi: 10.1111/cogs.12689

Martinet, A. (1952, April). Function, Structure, and Sound Change. WORD, *8*(1), 1–32. doi: 10.1080/00437956.1952.11659416

Morin, O., Winters, J., Müller, T. F., Morisseau, T., Etter, C., & Greenhill, S. J. (2018, July). What smartphone apps may contribute to language evolution research. *Journal of Language Evolution*, *3*(2), 91–93. doi: 10.1093/jole/lzy005

O'Donnell, T. J. (2015). *Productivity and Reuse in Language: A Theory of Linguistic Computation and Storage*. The MIT Press. doi: 10.7551/mitpress/9780262028844.001.0001

Plotkin, J. B., & Nowak, M. A. (2000, July). Language Evolution and Information Theory. *Journal of Theoretical Biology*, *205*(1), 147–159. doi: 10.1006/jtbi.2000.2053

Roberts, G., Lewandowski, J., & Galantucci, B. (2015, August). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66. doi: 10.1016/j.cognition.2015.04.001

Senghas, A., & Kita, S. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. , *305*, 5.

Smith, K., Kirby, S., & Brighton, H. (2003, October). Iterated Learning: A Framework for the Emergence of Language. *Artificial Life*, *9*(4), 371–386. doi: 10.1162/106454603322694825

Smith, K., Tamariz, M., & Kirby, S. (2013). Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35).

Spike, M., Stadler, K., Kirby, S., & Smith, K. (2017, April). Minimal Requirements for the Emergence of Learned Signaling. *Cognitive Science*, *41*(3), 623–658. doi: 10.1111/cogs.12351

Verhoef, T., Kirby, S., & de Boer, B. (2014, March). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57–68. doi: 10.1016/j.wocn.2014.02.005

Verhoef, T., Kirby, S., & de Boer, B. (2016, November). Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science*, *40*(8), 1969–1994. doi: 10.1111/cogs.12326

Yule, G. (2022). *The study of language*. Cambridge university press.

Zaslavsky, N., Garvin, K., Kemp, C., Tishby, N., & Regier, T. (2021, November). *The evolution of color naming reflects pressure for efficiency: Evidence from the recent past* (Preprint). Animal Behavior and Cognition. doi: 10.1101/2021.11.03.467047

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018, July). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942. doi: 10.1073/pnas.1800521115

Zipf, G. K. (1949). Human behavior and the principle of least effort.

Zuidema, W., & de Boer, B. (2018, June). The evolution of combinatorial structure in language. *Current Opinion in Behavioral Sciences*, *21*, 138–144. doi: 10.1016/j.cobeha.2018.04.011