

## Overview

**What we know:** Semantic systems reflect functional pressure for efficiency (e.g. Kemp et al. 2018; Zaslavsky et al. 2018).

**What we don't know:** What cultural process produces this pattern?

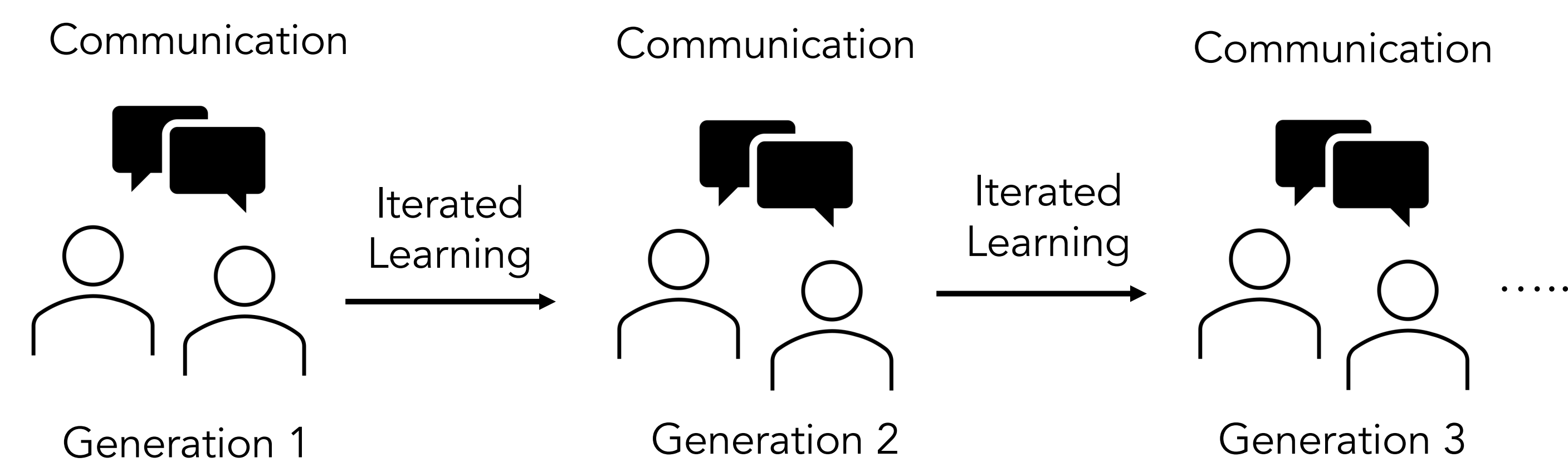
**Our contribution:** Human-like and efficient color naming systems emerge from a cultural evolutionary process that combines **iterated learning** and **communication**. Iterated learning alone, and communication alone, do not yield the same outcome as clearly.

## Semantic systems and cultural evolution

- Human color naming systems are efficient in the Information Bottleneck (IB) sense (Zaslavsky et al. 2018).
- How do semantic systems become efficient? Two proposed processes: **iterated learning** (Levinson 2012; J. Xu et al. 2013; Carstensen et al. 2015) and **communication** (e.g. Carr et al. 2020; Kågeback et al. 2020).
- Communicating artificial agents trained under reinforcement learning (RL) converge to IB-efficient color naming systems - but these systems are sometimes unlike human ones (Chaabouni et al. 2021).
- Communicating agents trained under an objective based on the IB objective converge to color naming systems that are both efficient and human-like (Tucker et al. 2022).

Does there exist a cultural evolutionary process that leads to IB-efficient color naming systems that are similar to human systems, but that is not directly based on the IB objective?

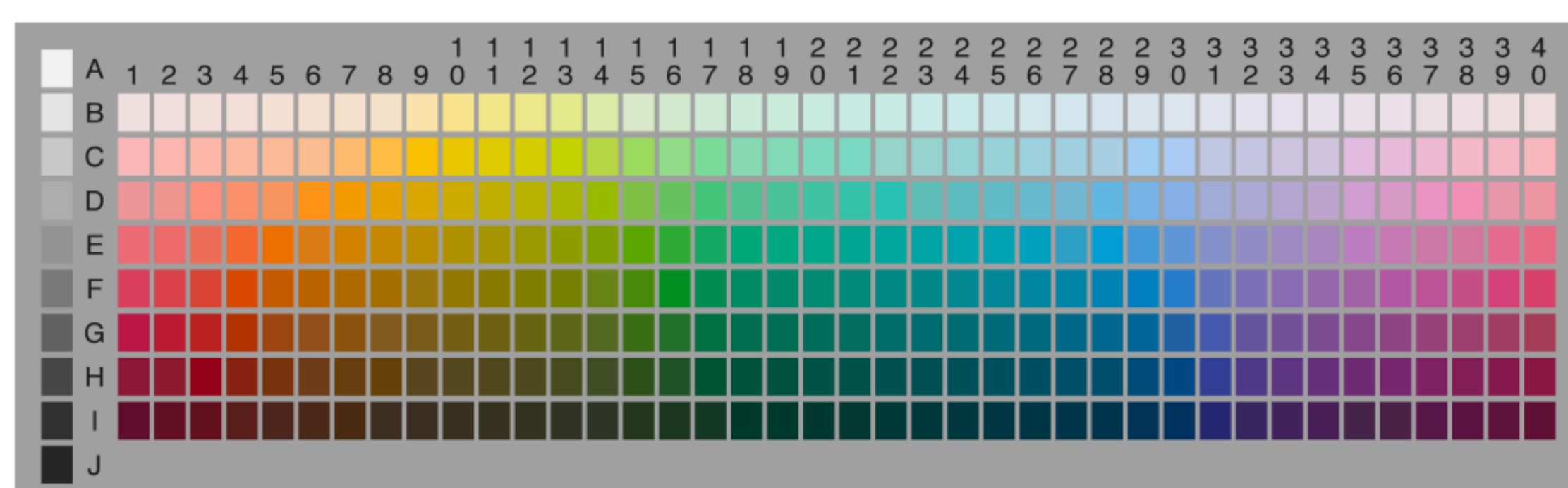
## Iterated learning with communication



Cultural evolution via the Neural Iterated Learning (NIL) algorithm (Ren et al. 2020):

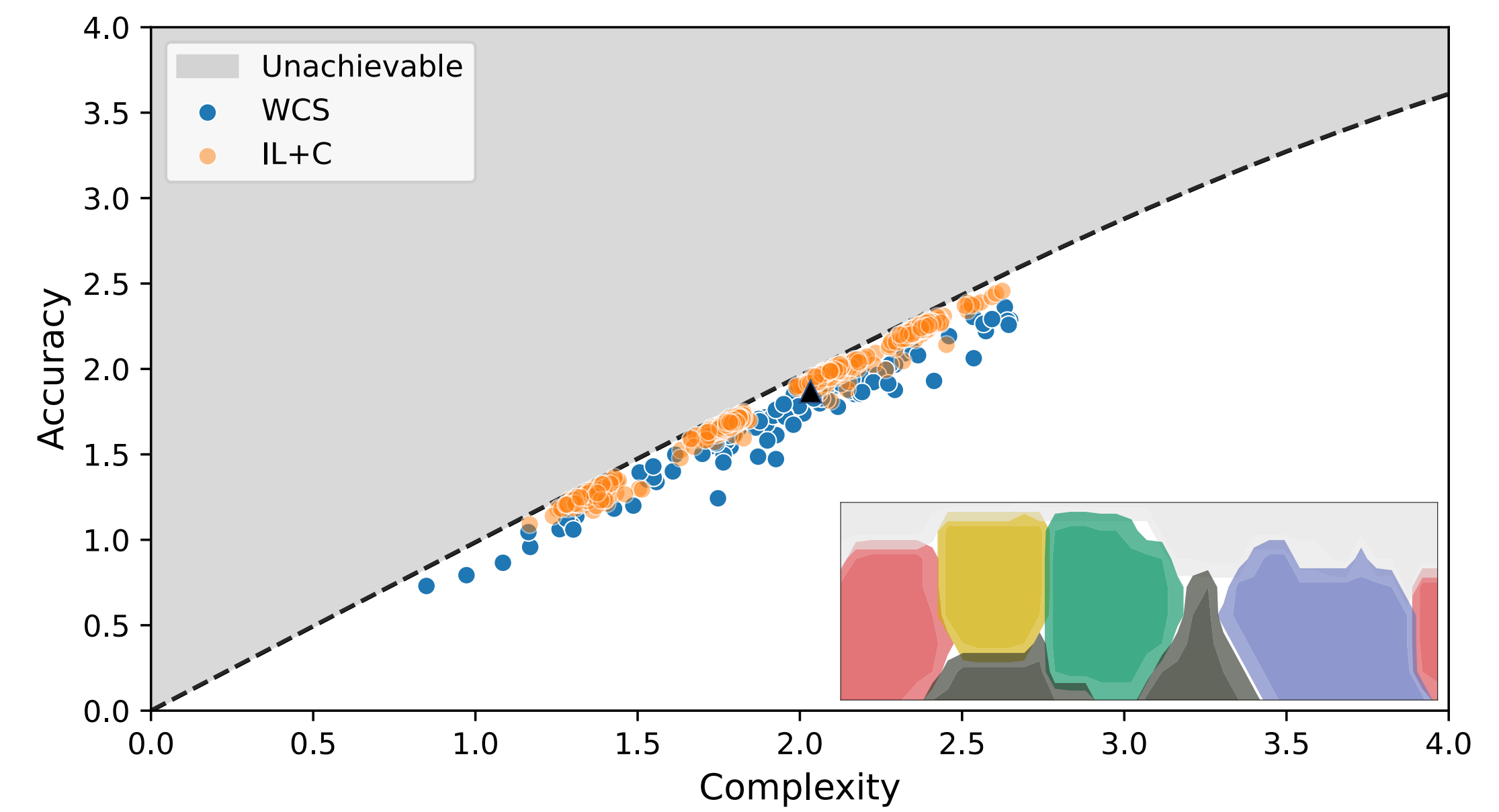
- Iterated learning (IL): each generation learns from samples drawn from the previous generation.
- Communication (C): Agents within a generation communicate, and update their behavior through reinforcement learning (Kågeback et al. 2020).

Assessed in the domain of color:

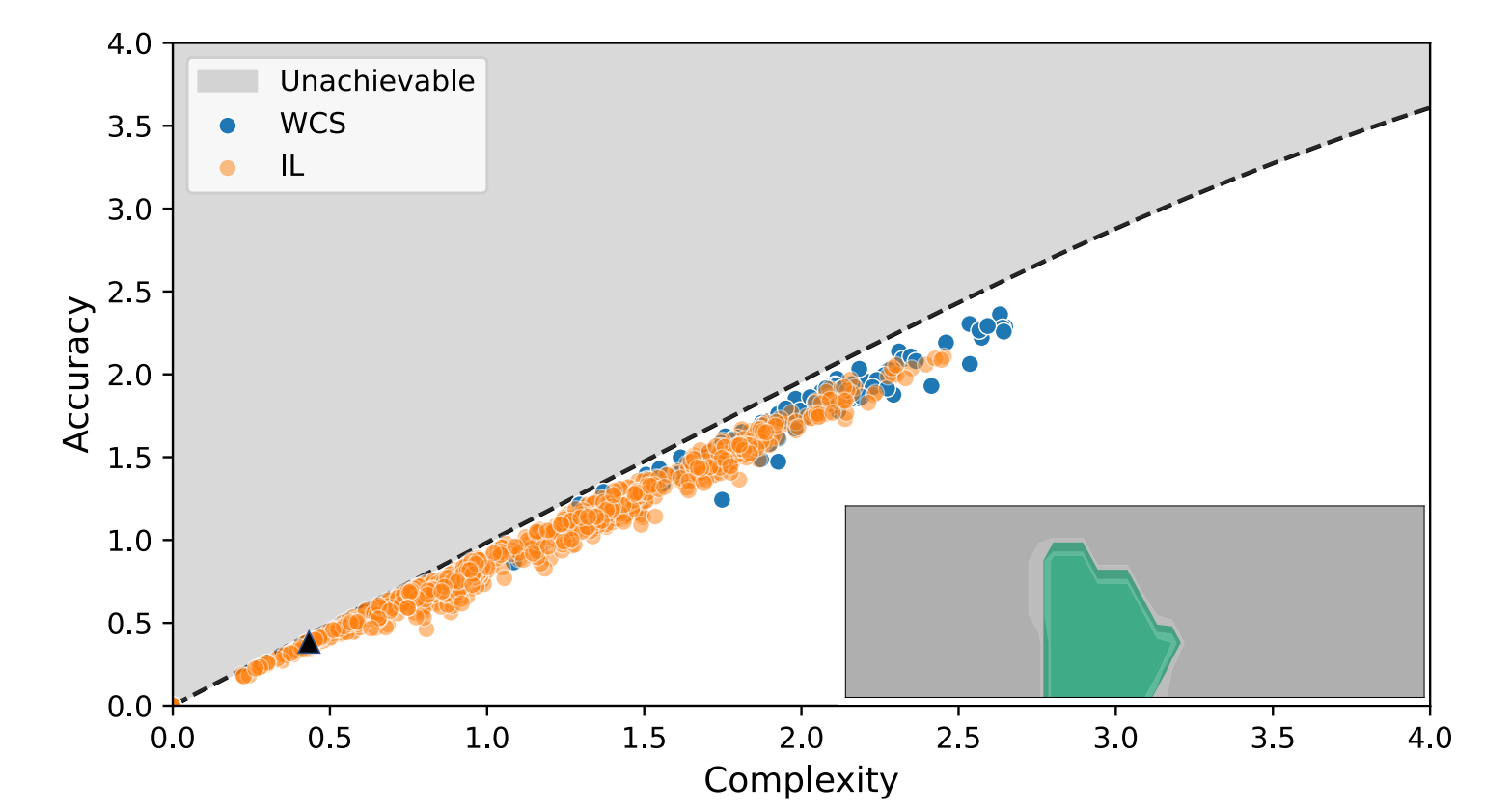
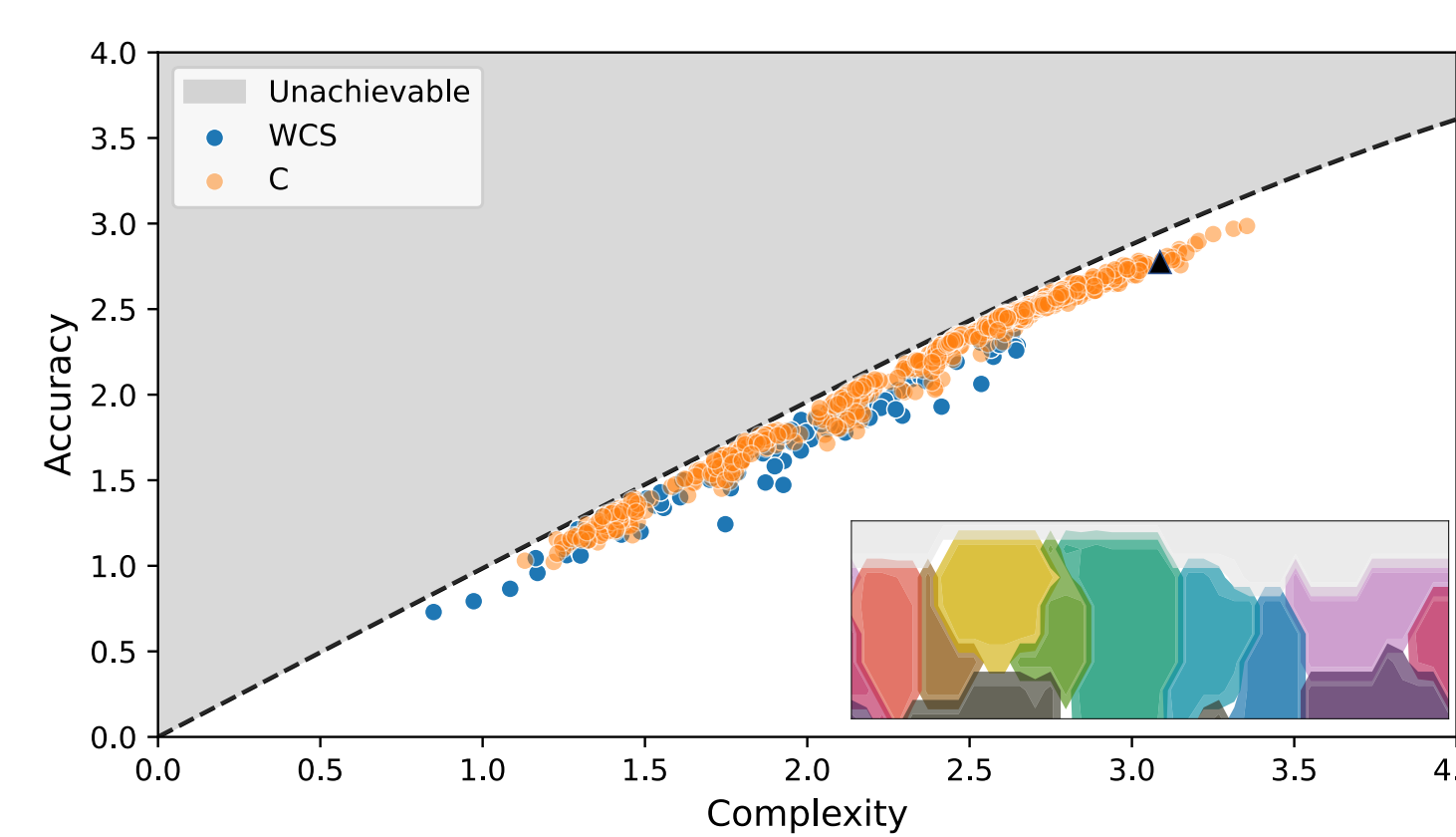


and results compared with human color naming systems of the World Color Survey (WCS; Cook et al. 2005).

## Results

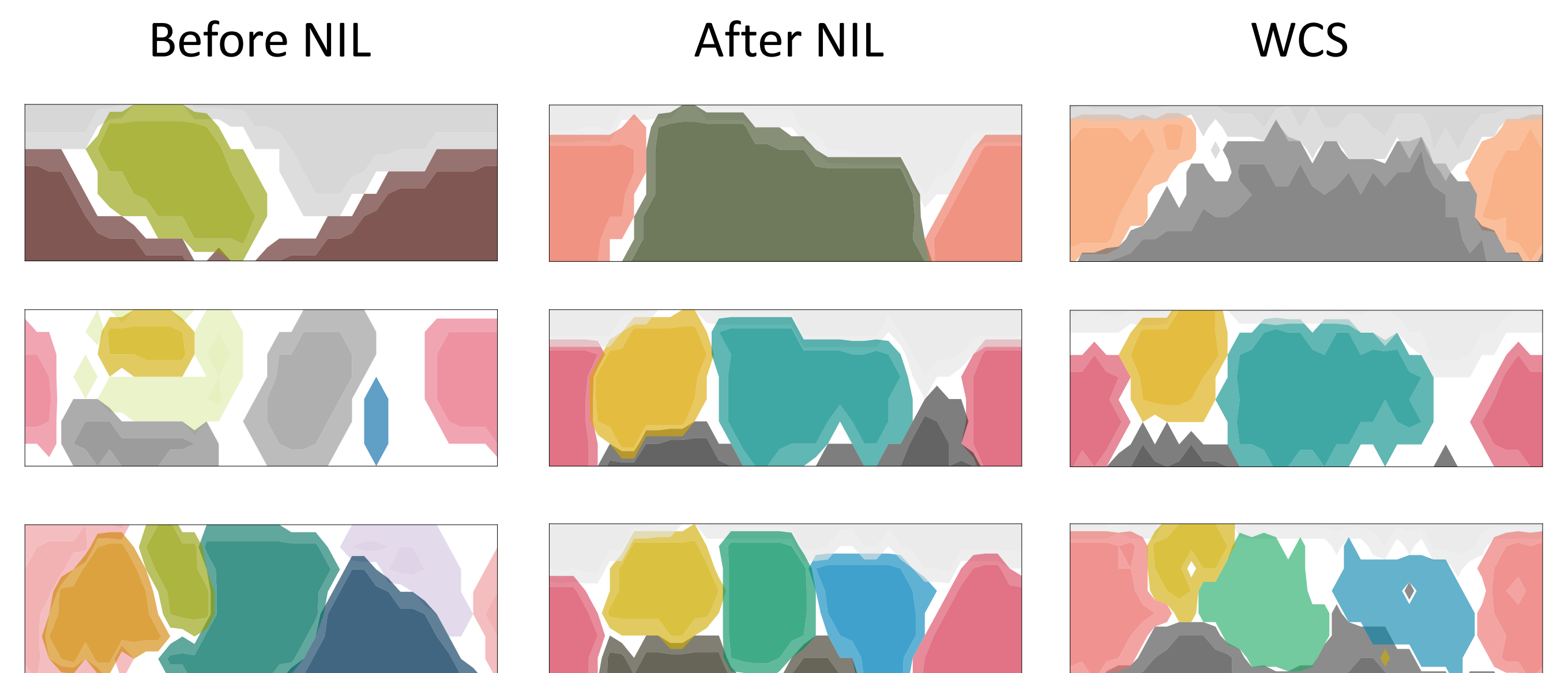


Iterated learning with communication (IL+C; i.e. standard NIL) generates color naming systems that are highly efficient in the IB sense, similar to those of the WCS, and in the same complexity and accuracy range as those of the WCS.



- Communication alone (C, left panel) often generates systems that are more complex than any WCS system.
- Iterated learning alone (IL, right panel) often generates systems that are simpler than any WCS system.

## NIL makes efficient systems more human-like



## Conclusions

Iterated learning combined with communication leads to IB-efficient color naming systems that are similar to human color naming systems.

Neither iterated learning alone, nor communication alone, accounts for human color naming as well.

Open question: What parts of this account are critical for the emergence of human-like color systems and what parts can be replaced with other processes?

## References

- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences of the United States of America*
- Ren, Y., Guo, S., Labeau, M., Cohen, S. B., & Kirby, S. (2020). Compositional languages emerge in a neural iterated learning model. In *International Conference on Learning Representations*.
- Carstensen, A., Xu, J., Smith, C. T., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.
- Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2020). Simplicity and informativeness in semantic category systems. *Cognition*, 202.
- Kågeback, M., Carlsson, E., Dubhashi, D., & Sayeed, A. (2020). A reinforcement-learning approach to efficient communication. *PLoS ONE*, 15(7)
- Tucker, M., Levy, R. P., Shah, J., & Zaslavsky, N. (2022). Trading off utility, informativeness, and complexity in emergent communication. *Advances in Neural Information Processing Systems*.
- Xu, Jing and Dowman, Mike and Griffiths, Thomas L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*.
- Stephen C. Levinson. (2012). Kinship and Human Thought. *Science*.
- Kemp, Charles and Xu, Yang and Regier, Terry. (2018). Semantic Typology and Efficient Communication.
- Chaabouni, Rahma and Kharitonov, Eugene and Dupoux, Emmanuel and Baroni, Marco. (2021). Communicating artificial neural networks develop efficient color-naming systems.
- Richard S. Cook and Paul Kay and Terry Regier. (2005). The World Color Survey Database: History and use. *Handbook of Categorization in Cognitive Science*.