Harmonizing Program Induction with Rate-Distortion Theory

Hanqi Zhou[1,2,3,4] (hanqi.zhou@uni-tuebingen.de), **David G. Nagy**[1,2,4], **Charley M. Wu**[1,2,4]

[1] Human and Machine Cognition Lab, University of Tübingen, Tübingen, Germany
[2] Max Planck Institute for Biological Cybernetics [3] IMPRS-IS [4] Tübingen AI Center

Preprint:

Background: Program induction under constraints

"Programs" of thought

Human concept learning can be viewed as inferring *Probabilistic Programs*[1]

- Finding the right **programs** involves optimizing over a combinatorial hypothesis space, which may exceed **human cognitive limitations**

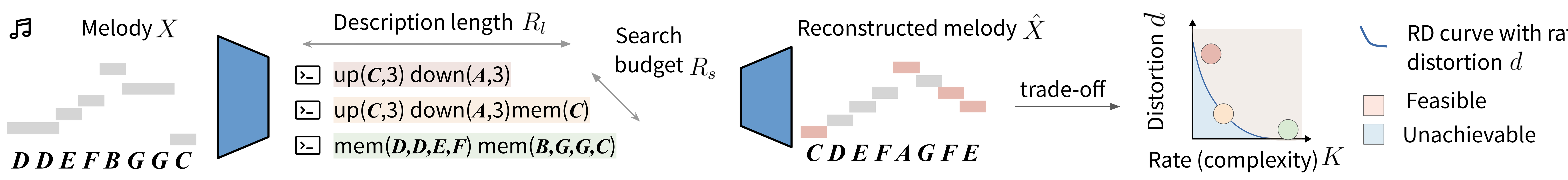
Resource rationality

Accounting for cognitive limitations is often modeled using information theory, specifically, *Rate-Distortion Theory* (RDT)[2]

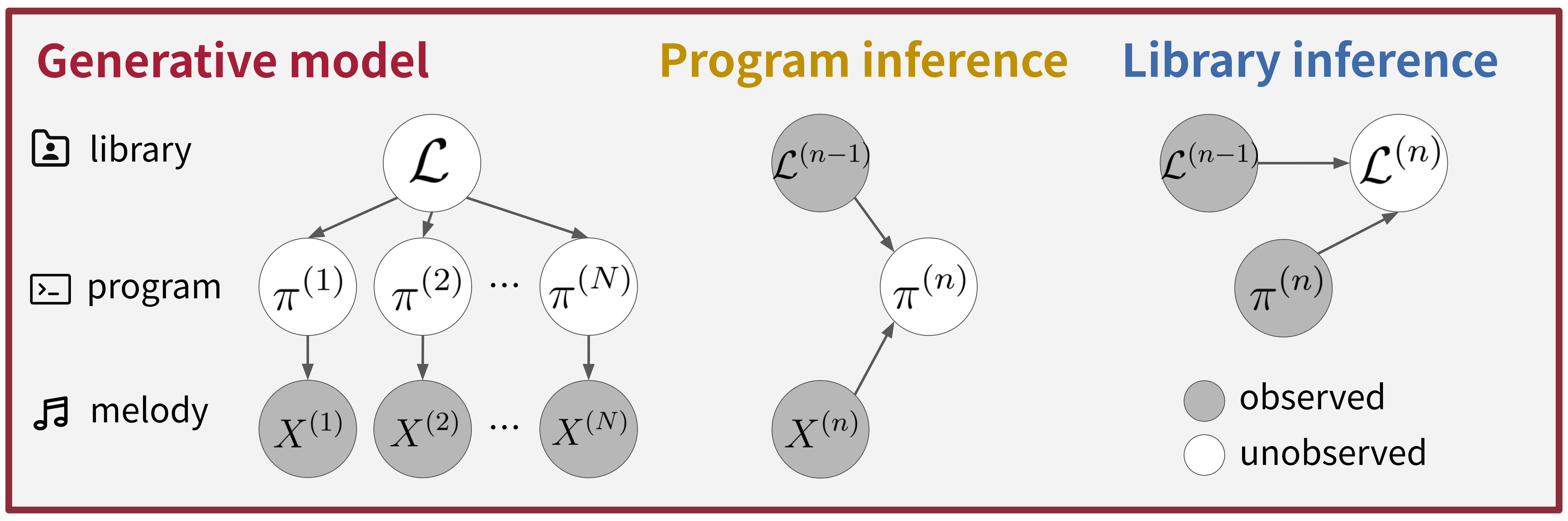
- Yet compression alone does not account for **computational costs** and how **learning changes the model**

We integrate program induction with RDT to explain how humans

- **Represent** knowledge as program induction under resource constraints incl. computational costs
- Continually **adapt** the repertoire of programs (i.e., compression model) to new observations



Simulation: Melody learning empowered by a shared library



Search for programs under constraints (within a melody):

- Description length R_l: An upper limit on the length of the programs
- Search budget R_s: An upper limit on how many programs are considered

Update the library to prefer useful programs (across melodies):

- Initial prior: proportional to **simplicity** (inverse length)
- Updated prior: proportional to **usage** based on Adaptor Grammar (AG)[3]

Connecting Bayes with RDT

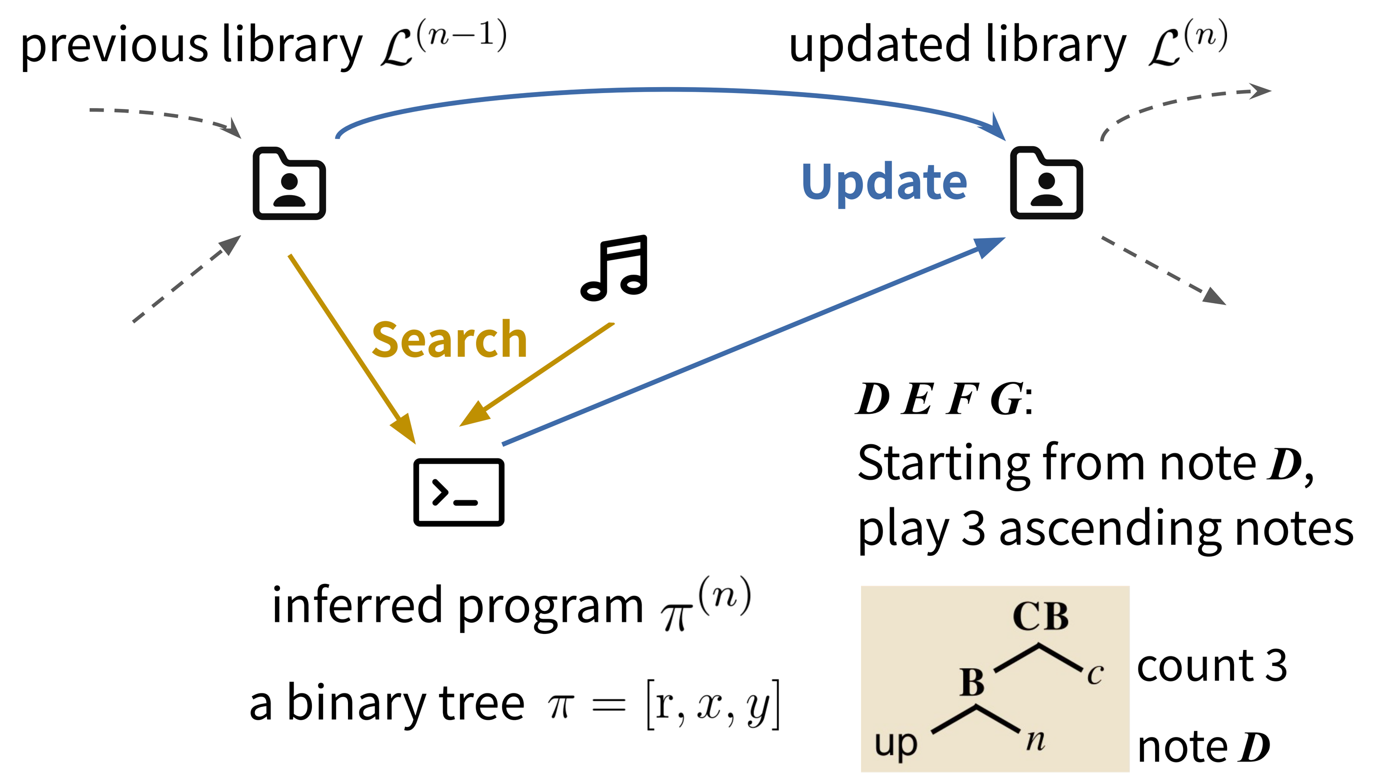
$$p(\pi \mid X) \propto p(X \mid \pi)\; p(\pi)$$

Likelihood $\propto$ Distortion^{-1}

Prior $\propto$ Description length^{-1}

Modeled via entropy coding

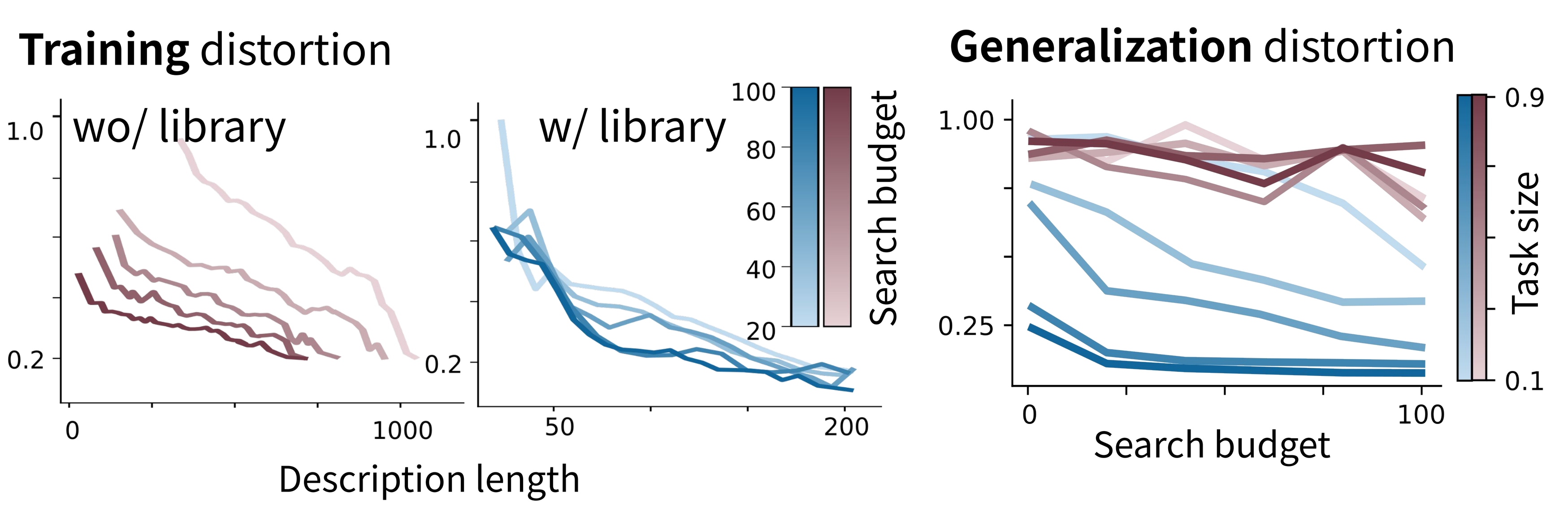
$$D_{\text{bounded}}(R) = \inf_{Q} D_Q, \text{ s.t. } K(\hat{X}) \leq R_l \wedge N_s \leq R_s$$



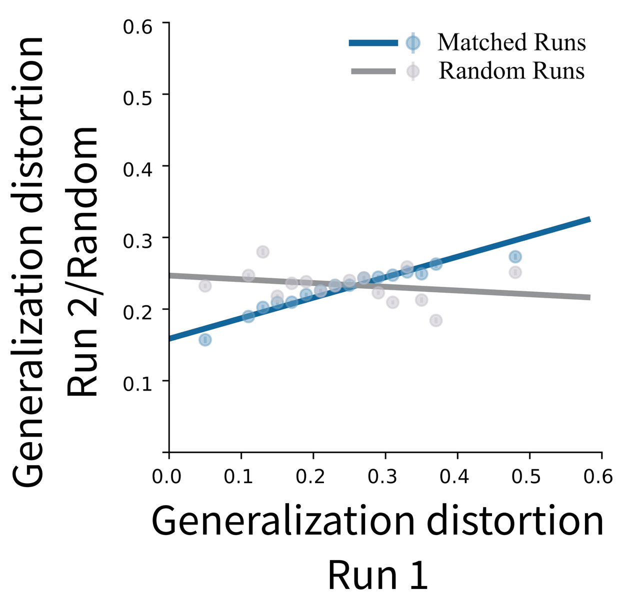
Results: Compression with curriculum effects

Modeling accumulated knowledge (program library) helps!

Lower distortion with the same resources, less sensitivity to the search budget, and better generalization to new data



But the library learning is path-dependent

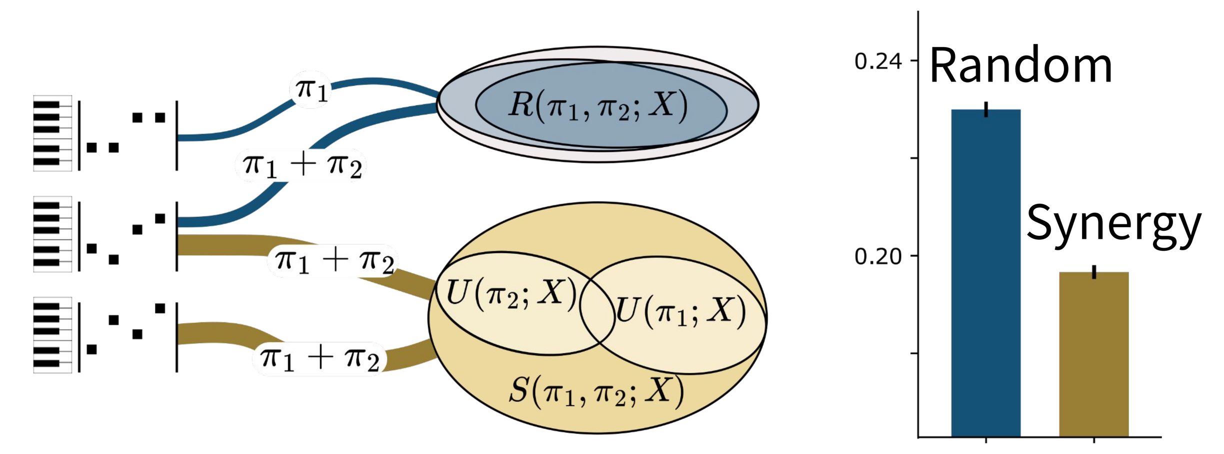


- Only ~3.19% of programs are the same under different curricula
- This cannot be accounted for by stochasticity of training alone (left)
- Humans also show path-dependence[4,5]

Can we leverage this to design better curricula *a priori*?

A synergistic curriculum building method using principles of partial information decomposition[6]:

$$S(\mathcal{L};X) = I(\pi_1, \pi_2; X) - R(\pi_1, \pi_2; X) - U(\pi_1; X) - U(\pi_2; X)$$



Future directions

- Fill the gap between these simulations and being able to model human learners on the same tasks
- Explore how different design choices, e.g., prior distributions, and task domains beyond melodies, would impact the results

References

[1] Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. Science, 350(6266), 1332-1338.
[2] Sims, C. R. (2016). Rate–distortion theory and human perception. Cognition, 152, 181-198.
[3] Liang, P., Jordan, M. I., & Klein, D. (2010, June). Learning programs: A hierarchical Bayesian approach. In ICML (Vol. 10, pp. 639-646).
[4] Zhao, B., Bramley, N. R., & Lucas, C. G. (2022). Powering up causal generalization: A model of human conceptual bootstrapping with adaptor grammars [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/7gvx9
[5] Dekker, R. B., Otto, F., Summerfield, C., Gershman, E. S. J., & Fiske, S. T. (2022). Curriculum learning for human compositional generalization. 119(41), 12.
[6] Proca, A. M., Rosas, F. E., Luppi, A. I., Bor, D., Crosby, M., & Mediano, P. A. (2024). Synergistic information supports modality integration and flexible learning in neural networks solving multiple tasks. PLOS Computational Biology, 20(6), e1012178.