Bootstrapping compositional generalization with cache-and-reuse

Bonan Zhao (b.zhao@ed.ac.uk)
Department of Psychology, University of Edinburgh

Christopher G. Lucas (c.lucas@ed.ac.uk)
School of Informatics, University of Edinburgh

Neil R. Bramley (neil.bramley@ed.ac.uk)
Department of Psychology, University of Edinburgh

Abstract

People effectively reuse previously learned concepts to construct more complex concepts, and sometimes this can lead to systematically different beliefs when the same evidence is processed in different orders. We model these phenomena with a novel Bayesian concept learning framework that incorporates adaptor grammars to enable a dynamic concept library that is enriched over time, allowing for caching and later reusing elements of earlier insights in a principled way. Our model accounts for unique curriculum-order and conceptual garden-pathing effects in compositional causal generalization that alternative models fail to capture: While people can successfully acquire a complex causal concept when they have an opportunity to cache a key sub-concept, simply reversing the presentation order of the same learning examples induces dramatic failures, and leads people to complex and ad hoc concepts. This work provides an explanation for why information selection alone is not enough to teach complex concepts, and offers a computational account of how past experiences shape future conceptual discoveries.

Introduction

People have a remarkable ability to develop rich and complex causal concepts despite our limited cognitive capacities (Griffiths, Lieder, & Goodman, 2015; Newell & Simon, 1972). We achieve this, partly, by effectively re-using, re-combining, or re-purposing existing knowledge (Gobet et al., 2001). This ability to bootstrap enables us to grow rich mental concepts incrementally that go beyond our limited cognitive resources—indeed, this is taken to be a cornerstone of cognitive development (Carey, 2004). Crucially, a model of conceptual bootstrapping posits learning not as optimal summarization of the environment, but as inherently and fundamentally path- and knowledge-dependent: Successful search for a complex causal concept is heavily reliant on having good, previously-learned abstractions (Dechter, Malmaud, Adams, & Tenenbaum, 2013; Gelpi, Prystawski, Lucas, & Buchsbaum, 2020).

Zhao, Bramley, and Lucas (2022) proposed a computational model of conceptual bootstrapping that attempts to formalize this crucial human learning ability in a Bayesian-symbolic concept learning framework. Drawing on combinatory logic and adaptor grammars (Liang, Jordan, & Klein,
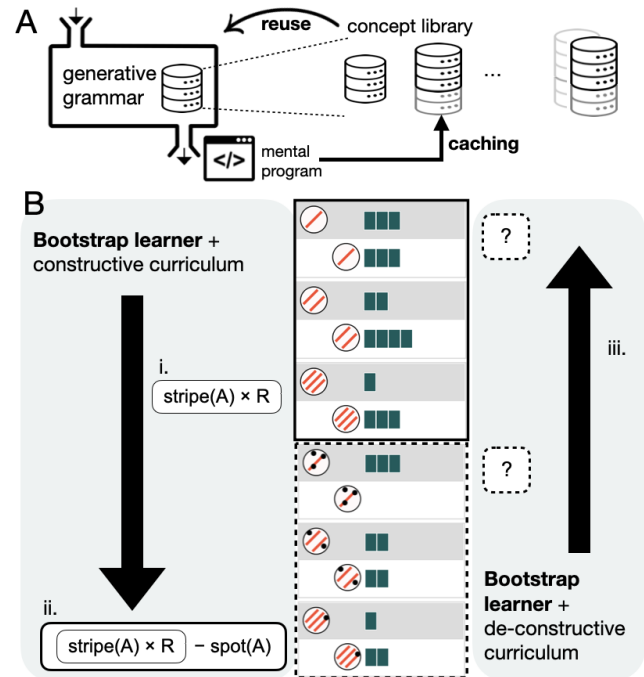

Figure 1: A. Visualization of the computational model. B. An example curriculum and corresponding model predictions.

2010), this model implements a dynamic concept library that can cache sub-concepts and therefore later reuse them to construct more complex concepts (Fig. 1A). This model predicts that processing the same information in different orders leads to different learned concepts, because of different sub-concepts acquired in the process of learning. We further test the validity of this model by conducting replicating and follow-up experiments, and comparing the original model with several alternative models.

Methods

Experiments

Zhao et al. (2022) tested causal concept learning using animated geometric objects called a causal agent (A), recipient (R), and result (R'). An agent object A has stripes and spots on it, and a recipient object R is composed of several segments (Fig. 1B, shaded panels). When A touches R, A can change the number of segments on R to change, bringing the result R' (Fig. 1B, white panels). Zhao et al. (2022) tested causal concept $R' = \mathtt{stripe}(A) \times R - \mathtt{spot}(A)$.
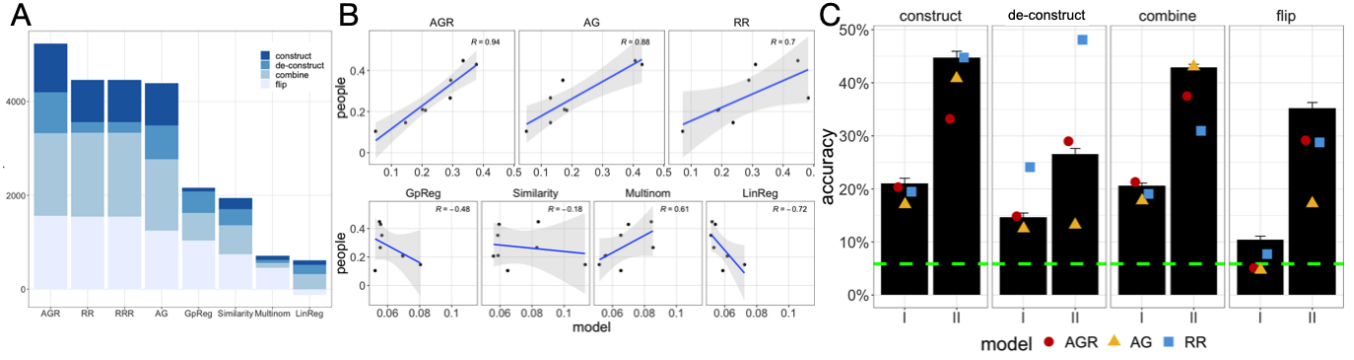
Figure 2: A. Model fit (total log likelihood) improvement over random baseline (y=0), log scale. B. Generalization accuracy per curriculum and phase. X-axis are model predictions, y-axis people's. C. Generalization accuracy between people (black bars) and four Bayesian-symbolic models.

Here, we replicated this experiment reported by swapping the causal roles attached to the stripe and spot features, i.e., $R' = \text{spot}(A) \times R - \text{stripe}(A)$.

Zhao et al. (2022) manipulated three curricula, each consisting of two phases. Curriculum *construct* presents $R' = \text{stripe}(A) \times R$ in Phase I (Fig. 1B, solid-border box), and then information about the compound concept in Phase II (Fig. 1B, dashed-border box). Curriculum *de-construct* reverses Phase I and II as in *construct*. Curriculum *combine* shares the same Phase I as in *construct*, but in Phase II kept $\text{stripe}(A) = 1$ throughout, making it ambiguous how $\text{stripe}(A) \times R$ and $\cdot - \text{spot}(A)$ should be combined. We further included a *flip* curriculum that swaps the two phases of *combine* to explicitly test how people do this. We tested both the causal function in Zhao et al. (2022) and its replication in these experiments.

Participants watched animated object interactions during training, and then made predictions on eight novel $\langle A, R \rangle$ in random order, once after Phase I and once after Phase II.

Models

We compared an extended version of the Adaptor Grammar model (AG) reported in Zhao et al. (2022) and several alternative models. The extended model is called AGR—Adaptor Grammar with Re-processing. This model mixes predictions $\hat{y}_\rightarrow$ from Phase I to II, and predictions $\hat{y}_\leftarrow$ from Phase II to I, with a weight parameter $\lambda \in [0, 1]$, getting a mixed prediction $\hat{y}_r \propto \lambda \cdot \hat{y}_\rightarrow + (1 - \lambda) \cdot \hat{y}_\leftarrow$. This is to account for the fact that people may be revisiting Phase I after seeing Phase II, as allowed by the experiment interface.

For alternative models, we first examined a "rational rules" model (RR) based on Goodman, Tenenbaum, Feldman, and Griffiths (2008), assuming the same conceptual primitives as the adaptor grammar models. Since we evaluate models using generalization predictions, we also implemented several sub-symbolic models capable of generalization but not explicit rule guesses: A similarity-based categorization model (Tversky, 1977), a linear regression model (LinReg), a multinomial regression model (Multinom) and a Gaussian process regression (GpReg) model with radial basis function kernels

one per feature.

We then made predictions about the novel object pairs using fitted models, and evaluated model predictions with their log-likelihood LL of producing participants' predictions. The baseline model is random selection (rand).

Results and Discussion

Since the feature-counterbalancing did not interfere with the main behavioral results, we collapsed across these in analysis. As shown in Fig. 2A, model AGR achieves the greatest improvement of fit over baseline, with the three Bayesian-symbolic models (AGR, AG, RR) easily outperforming similarity-based or regression models. With fitted model parameters, Fig. 2B plots generalization accuracy in each phase for each curriculum between model and people. Again, AGR best predicts people's performance across all cases, and the non-symbolic models fail to match people's predictions.

Fig. 2C plotted average accuracy achieved by human participants in black bars. The difference in Phase II accuracy between the *construct* and *de-construct* curricula revealed strong curriculum-order and conceptual garden-pathing effects. Model RR fails to reproduce such effects, because this model is likely to have figured ground truth after seeing all the data, even for the *de-construct* curriculum, and thus deviating from how people process phases of information. Model AG, on the other hand, is defeated by the learning trap as many people were, exhibiting no accuracy improvement in Phase II relative to Phase I. Model AGR mixes model AG with some re-processing, and is therefore able to capture participants' modest improvement in *de-construct* Phase II generalizations.

Overall, these results suggest that curriculum-order and conceptual garden-pathing effects exhibited by people can be explained as consequences of a cache-and-reuse mechanism expanding the reach of a bounded learning system. Critically, these phenomena cannot be explained by a standard Bayesian-symbolic model out of the box, or by familiar sub-symbolic categorization models, showcasing that a cache-and-reuse mechanism is central to human-like inductive inference to compositional concepts.

Acknowledgments

References

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59–68.

Dechter, E., Malmaud, J., Adams, R. P., & Tenenbaum, J. B. (2013). Bootstrap learning via modular concept discovery. In *Twenty-third international joint conference on artificial intelligence.*

Gelpi, R., Prystawski, B., Lucas, C. G., & Buchsbaum, D. (2020). Incremental hypothesis revision in causal reasoning across development. In *Proceedings of the 42th annual conference of the cognitive science society.*

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Liang, P., Jordan, M. I., & Klein, D. (2010). Learning programs: A hierarchical Bayesian approach. In *Proceedings of the 27th international conference on machine learning (icml-10)* (pp. 639–646).

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Zhao, B., Bramley, N. R., & Lucas, C. G. (2022). Powering up causal generalization: A model of human conceptual bootstrapping with adaptor grammars. In *Proceedings of the 44th annual meeting of the cognitive science society* (pp. 1819–1826).