Order Effects in One-shot Causal Generalization

Bonan Zhao[1] (b.zhao@ed.ac.uk) **Neil Bramley**[1] (neil.bramley@ed.ac.uk) (& in-process new work with **Chris Lucas**[2] (c.lucas@ed.ac.uk))
{[1]Department of Psychology, [2]School of Informatics}, The University of Edinburgh, Edinburgh, Scotland, EH8 9JZ UK

THE UNIVERSITY of EDINBURGH

Overview

- People make systematic causal generalizations from a single example
- Order of generalizations matters:
  - When first generalizing to objects *similar* to the example people draw strong systematic conclusions that carry over to objects *dissimilar* to the example (*near-transfer* condition), however
  - …when first generalizing to objects *dissimilar* to learning case people draw diverse conclusions that again carry over to objects *similar* to the example (*far-transfer* condition).
- Pattern can be captured by a greedy nonparametric causal categorization model

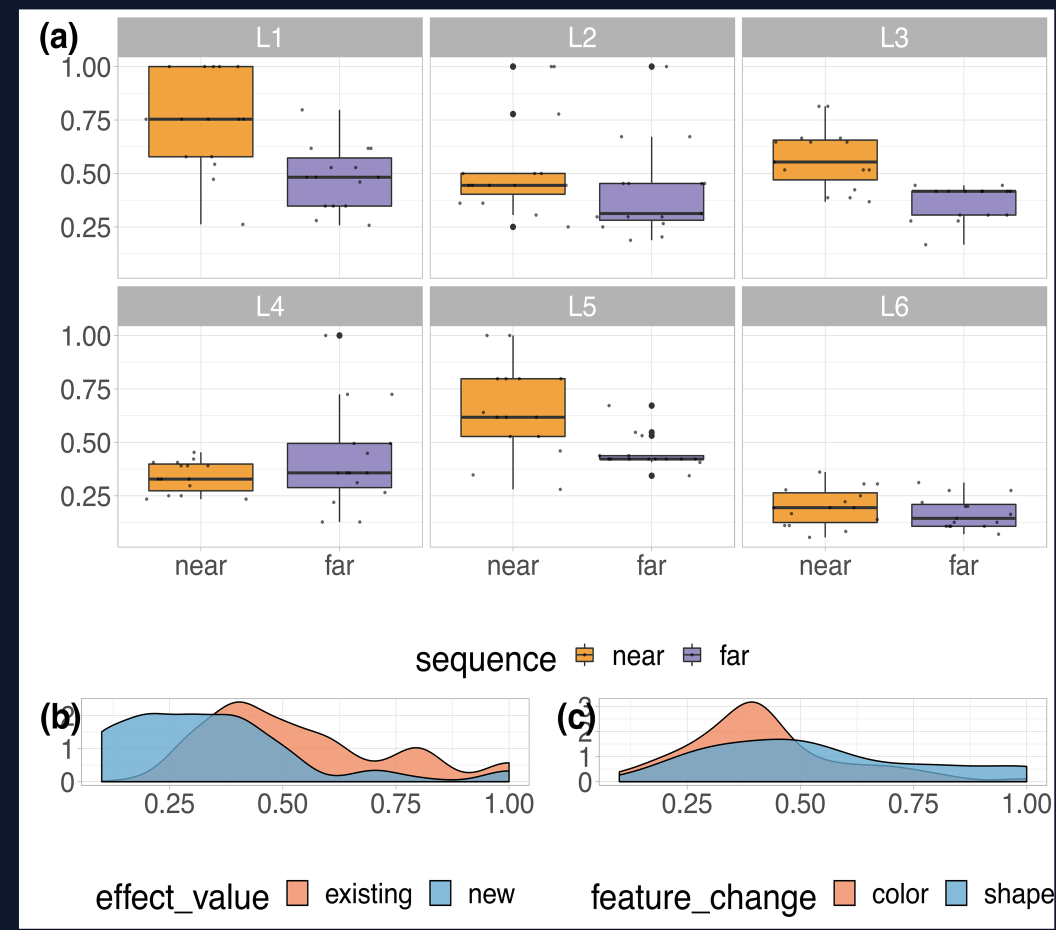
Experiment

- 6 learning scenes
- 15 generalization trials per learning scene
- 2 presentation orders of trials
- 120 participants on Mturk

Behavioral Results

***η* measures "agreement":**

- In all cases, people agreed far above chance
- Near-transfer condition produced higher agreement overall
- Colors and shapes were generalized to different extents - people agreed more for shape-related changes



Computational Models

Normative Model

- Hypothesis space: causal "rules" generated by a Probabilistic Context Free Grammar (PCFG)
- + Causal object categories capturing provenance of causal functions over types of Agent (cause) and Recipient (effect) objects modeled with a Dirichlet Process

Process Model

- Learner's generate single causal category & function assignment for each generalization and commit to it as "data"

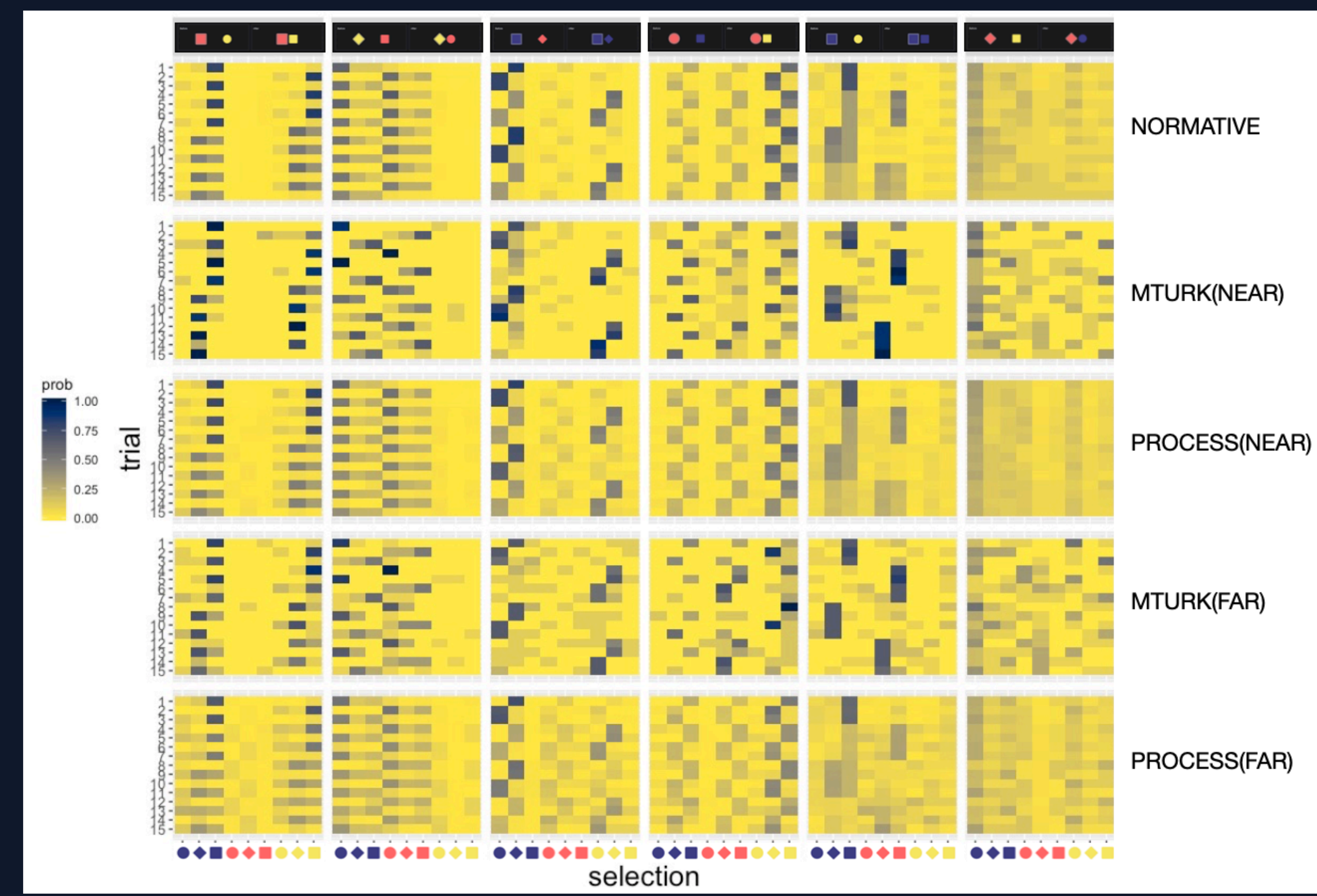
Table 1: λ-abstraction Based Probabilistic Grammar.

Productions			
Start	$S \rightarrow$	$\lambda_{\phi_i} : A, \Phi$	
Bind additional	$A \rightarrow$	B	$B \wedge S$
Relation	$B \rightarrow$	$(r' = C)$	$(r' \neq C)$
Reference	$C \rightarrow$	$D1$	$D2$
Relative ref.	$D1 \rightarrow$	a_{ϕ_i}	r_{ϕ_i}
Absolute ref.	$D2 \rightarrow$	value[a]	

[a]: Sampled uniformly from the support of feature ϕ_i.

Data: $D_L = \{d_0\}, D_G = \{d_1, \dots, d_n\}$
Result: $N_G = \{n_1, \dots, n_n\}$
Assign d_0 to category Cat_1;
Sample f_{Cat_1} for Cat_1 from the posterior;
for $d_i \in D_G$ **do**
  if d_i *belongs to an existing category* Cat_j **then**
    $n_i = f_{Cat_j}(d_i)$;
  else
    Create Cat_{new};
    Sample $f_{Cat_{new}}$ from the prior;
    $n_i = f_{Cat_{new}}(d_i)$;
  end
end

Model Fits



Model	Likelihood	BIC	R^2
Random baseline	-3955	7910	
Normative model*	-2687	5389	.63
Process model*	-2642	5299	.65

*With initial parameters $\mu = 0.1, \alpha = 0.1$.

Note: Model fitting with participant data is still on-going; results here are with initial parameters.

References

- Bramley, N. R., Rothe, A., Tenenbaum, J., Xu, F., & Gureckis, T. (2018). Grounding compositional hypothesis generation in specific instances. In *Proceedings of the 40th annual conference of the cognitive science society*.
- 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., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review*, *116*(4), 661. Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program in- duction. *Science*, *350*(6266), 1332–1338.
- Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*(1), 113–147.

<- Try the experiment

Read manuscript ->

Experiment URL http://bramleylab.ppls.ed.ac.uk/experiments/bnz/magic_stones/index.html
Manuscript URL https://github.com/zhaobn/magic_stones/blob/master/drafts/zhao2020cgw.pdf