The Role of Variability in Learning Generalization: A Computational Modeling Approach

Thomas Gorman

Learning Generalization

- Learning is often specific
- Longstanding scientific interest in how to improve generalization or transfer

Variability

- Variation during training linked to improved transfer in numerous domains
- What does “variability mean in the context of learning interventions?”

Types of Variability

- The number of unique items/problems experienced
- How spread out examples are

My Dissertation Focus

- Number of unique examples
- Visuomotor skill learning and function learning
- Addressing methodological shortcomings of previous work
- Adapting cognitive models from other domains to account for results

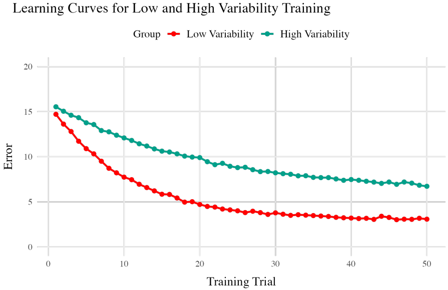
Common Experimental Manipulations

- Varied vs. Constant

Common Empirical Patterns

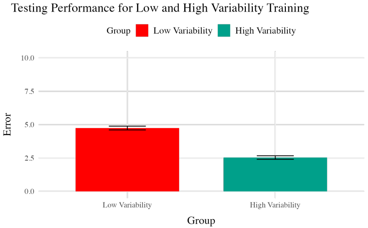
Training

- Both training conditions complete the same number of training trials.
- Varied group has worse training performance.



Testing

- Tested from novel conditions.
- Varied group has better test performance



But also plenty of contradictory results and complications

- Cases where varied training doesn't benefit generalization
- Cases where more training variation results in worse outcomes
- Cases where the influence of variation interacts with some other factor
  - difficulty
  - prior knowledge
  - Frequency effects, or amount of training/learning before testing

Theoretical Frameworks

- Schema Theory (Schmidt, 1975)
- Desirable Difficulties Framework (Bjork & Bjork, 2011)
- Challenge Point Framework (Guadagnoli & Lee, 2004)

Schmidt (1975) Bjork & Bjork (2011) Guadagnoli & Lee (2004)

Overview of Current Work

Project 1[^1]

- Visuomotor projectile launching task
- two experiments
- Beneficial effect of variability
- Instance-based similarity model

Project 2

- Visuomotor extrapolation task
- Three experiments
- Effect of variability is null or negative
- Connectionist model (ALM) and hybrid associative & rule model (EXAM)

–

Project 1 - Issues with previous research

- Assumptions about what is encoded
- Assumptions about the formation of abstractions
- Aggregation issues and similarity confounds

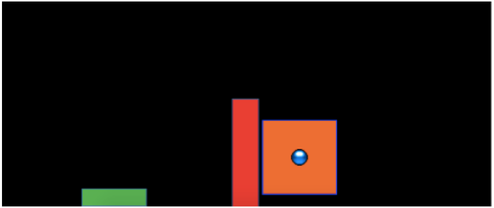
designs that seemingly overcome these issues

- Kerr & Booth 1978

Experiment 1

- Conceptual replication of Kerr & Booth design
- Testing classic pattern as well

Hit The Target Task

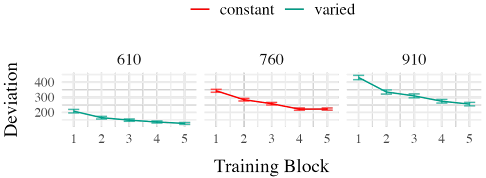


- **Training Stage** - 200 training trials with feedback. Constant groups trains from single position. Varied group practices from two positions.
- **Transfer Stage** - All subjects tested from both positions they were trained, and the positions trained by other group
- **Data recorded** - For every throw, recorded the X velocity and Y velocity of ball at release

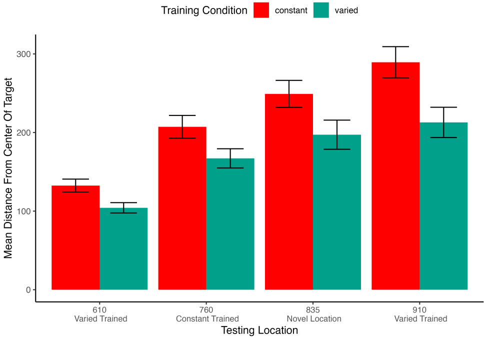
Project 1 - Experiment 1 Results

Experiment 1

Training



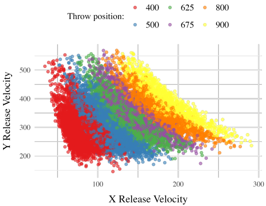
Testing

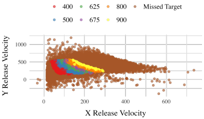


Project 1 - Discussion

- Varied training group's superior performance
- Computational modeling results with IGAS
- Theoretical implications

Project 1 Computational Model





Model Steps

- each of the 6 positions has an empirical solution space
- Compute the summed similarity between training throws, and solution spaces
  - separately for each participant, and each of the testing positions
- We now have a measure of how similar training behavior was to testing solutions

Model Definition

- $d_{i,j} = \sqrt{(x_{Train_i} - x_{Solution_j})^2 + (y_{T}$
- $Similarity_{I,J} =$

Project 2 - Variability and Extrapolation in a Function Learning Task

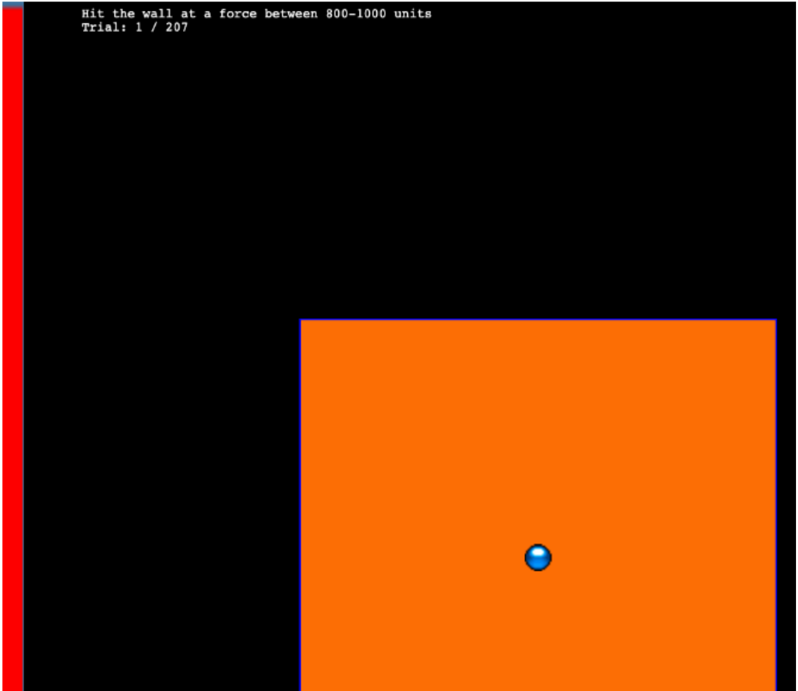
- Influence of varied practice in a function learning task
- Experiments 1, 2, and 3:
  - Training regimes and testing conditions
  - Learning, discrimination, and extrapolation performance metrics

Project 2 - Questions and Goals

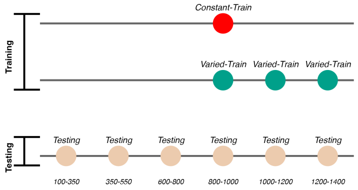
Empirical - Design a task-space large enough to assess multiple degrees of extrapolation - Compare varied and constant generalization from several distinct distances from their nearest training condition

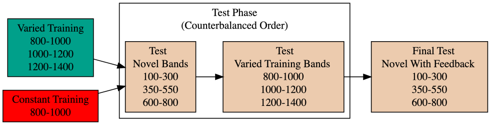
Model-based - If variation does influence extrapolation, can an associative learning model with similarity-based activation provide a good account? - Can our modelling framework simultaneously account for both training and testing data? - Accounting for the full distribution of responses

Hit The Wall Task



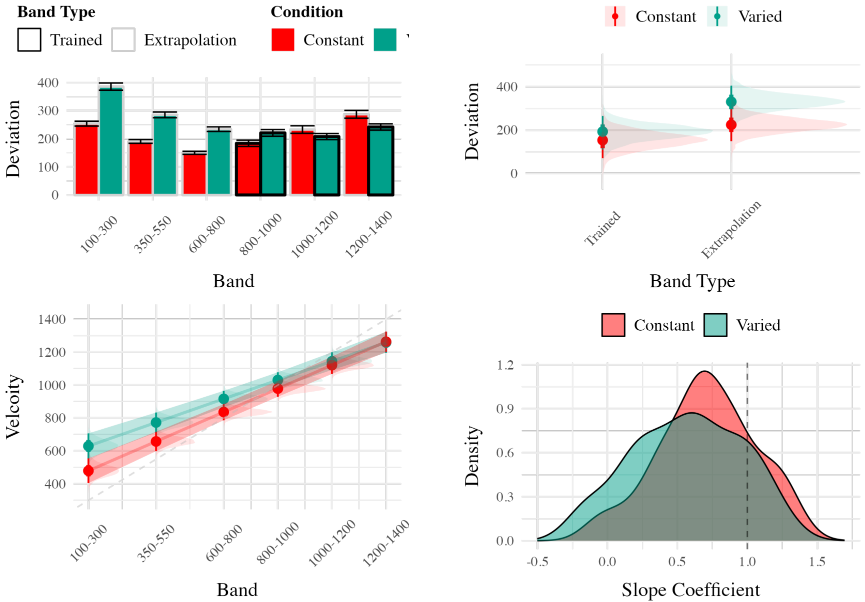
Project 2 - Experiment 1 Design





- 156 participants included in final analysis
- Varied group trains from 3 "velocity bands", constant group from 1

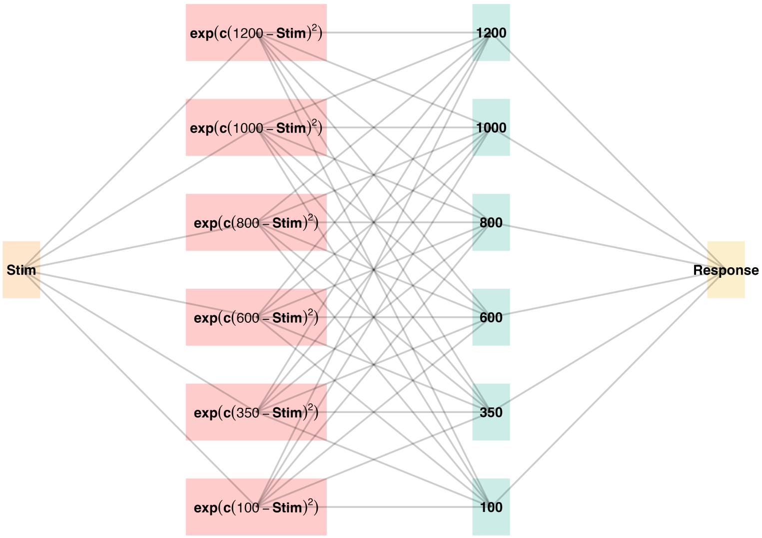
Project 2 - Experiment 1 Results



Project 2 - Experiment 2 Design

Project 2 - Computational Modeling

Project 2 - ALM



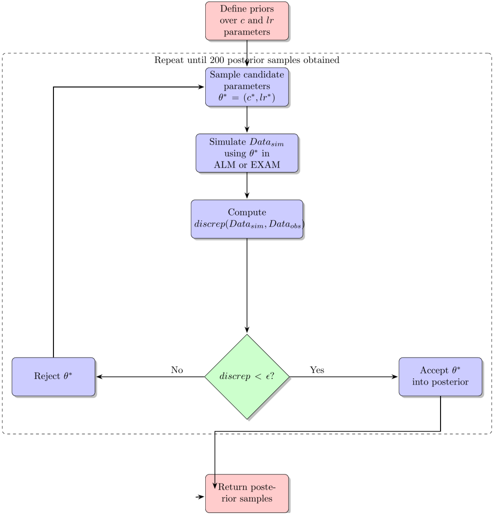
	ALM Response Generation	
Input Activation	$a_i(X) = \frac{e^{-c(X-X_i)^2}}{\sum_{k=1}^{M} e^{-c(X-X_k)^2}}$	Input nodes activate as a function of Gaussian similarity to stimulus
Output Activation	$O_j(X) = \sum_{k=1}^{M} w_{ji} \cdot a_i(X)$	Output unit O_j activation is the weighted sum of input activations and association weights
Output Probability	$P[Y_j \mid X] = \frac{O_j(X)}{\sum_{k=1}^{M} O_k(X)}$	The response, Y_j probabilites computed via Luce's choice rule
Mean Output	$m(X) = \sum_{j=1}^{L} Y_j \cdot \frac{O_j(x)}{\sum^{M} O_k(X)}$	Weighted average of probabilities

EXAM

	EXAM Response Generation	
Instance Retrieval	$P[X_i\|X] = \frac{a_i(X)}{\sum_{k=1}^{M} a_k(X)}$	Novel test stimulus X activates input nodes X_i
Slope Computation	$S = \frac{m(X_1)-m(X_2)}{X_1-X_2}$	Slope value, S computed from nearest training instances
Response	$E[Y\|X_i] = m(X_i) + S \cdot [X - X_i]$	Final EXAM response is the ALM response for the nearest training stimulus, $m(X_i)$, adjusted by local slope S.

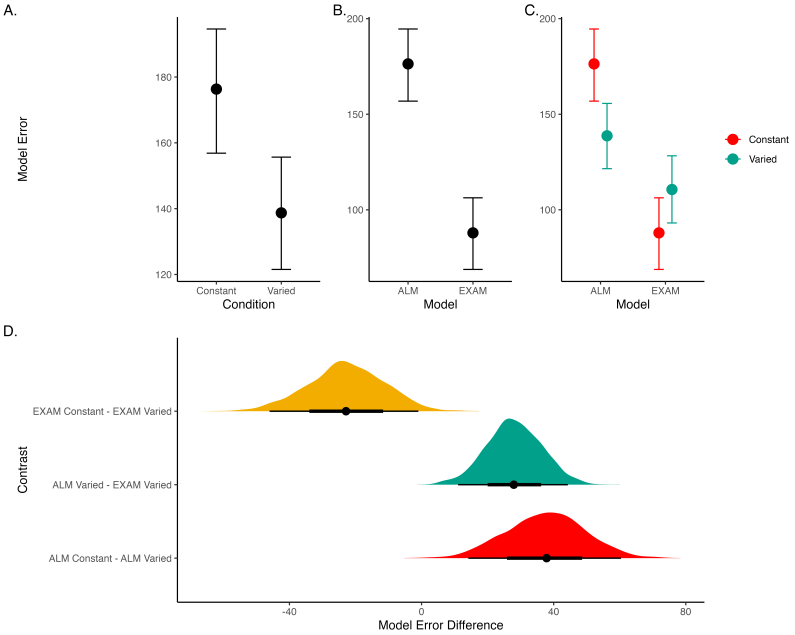
Project 2 - Model Fitting Procedure

Approximate Bayesian Computation



- simulation based approach
- approximate likelihood
- uncertainty in parameter values
- full distribution of plausible model predictions for each participant

Modelling Results



General Discussion

- Compare HTT and HTW
- Empirical findings summary
- Computational modeling contributions

Comparison

Dimension	HTT (Project 1)	HTW (Project 2)
Task Description	Projectile launching to hit a target	Projectile launching to hit wall at a specific velocity
Task Complexity	More complex parabolic trajectory, both x and y velocities relevant to outcome	Simpler 1D mapping of force to outcome. Only x velocity is relevant.
Task Space	More complex: xy velocity combinations closer to the solution space may still result in worse feedback due to striking the barrier.	Simpler: smooth, linear mapping between velocity and feedback.
Perceptual salience of Varied Conditions	Varied conditions (# of throwing distances) are perceptually distinct, i.e. salient differences in distance between	Varied conditions (# of velocity bands) are less salient - only difference is the numeral displayed on

Conclusion

References

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, *2*, 59–68.

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning. *Journal of Motor Behavior*, *36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770