Full Dissertation2

The Role of Variability in Learning Transfer: A Similarity-Based Computational Approach

Thomas Gorman

Table of contents

# Abstract

The past century of research on human learning has produced ample evidence that although learners can improve at almost any task, such improvements are often specific to the trained task, with unreliable or even nonexistent transfer or generalization to novel tasks or conditions (Barnett & Ceci, 2002; Detterman, 1993). Such generalization challenges are of noteworthy practical relevance, given that educators, trainers, and rehabilitators typically intend for their students to be able to apply what they have learned to new situations. It is therefore important to better understand the factors that influence generalization, and to develop cognitive models that can predict when generalization is likely to occur. Such characteristics have included training difficulty, spacing, temporal order, feedback schedules, and the primary focus of the current work - the variability of training examples.

Of course, the relationship between training variability and transfer is unlikely to be a simple function wherein increased variation is always beneficial. Numerous studies have found null, or in some cases negative effects of training variation (DeLosh et al., 1997; Sinkeviciute et al., 2019; Van Rossum, 1990; Wrisberg et al., 1987), and many more have suggested that the benefits of variability may depend on additional factors such as prior task experience, the order of training trials, or the type of transfer being measured (Berniker et al., 2014; Braithwaite & Goldstone, 2015; Hahn et al., 2005; Lavan et al., 2019; North et al., 2019; Sadakata & McQueen, 2014; Zaman et al., 2021).

In an example of a more complex influence of training variation, (Braithwaite & Goldstone, 2015) trained participants on example problems involving the concept of sampling with replacement (SWR). Training consisted of examples that were either highly similar in their semantic context (e.g. all involving people selecting objects) or in which the surface features were varied between examples (e.g. people choosing objects AND objects selected in a sequence). The experimenters also surveyed how much prior knowledge each participant had with SWR. They found that whether variation was beneficial depended on the prior knowledge of the participants – such that participants with some prior knowledge benefited from varied training, whereas participants with minimal prior knowledge performed better after training with similar examples. The authors hypothesized that in order to benefit from varied examples, participants must be able to detect the structure common to the diverse examples, and that participants with prior knowledge are more likely to be sensitive to such structure, and thus to benefit from varied training. To test this hypothesis more directly, the authors conducted a 2nd experiment, wherein they controlled prior knowledge by exposing some subjects to a short graphical or verbal pre-training lesson, designed to increase sensitivity to the training examples. Consistent with their hypothesis, participants exposed to the structural sensitivity pre-training benefited more from varied training than the controls participants who benefited more from training with similar examples. Interactions between prior experience and the influence of varied training have also been observed in sensorimotor learning (Del Rey et al., 1982; Guadagnoli et al., 1999). Del Rey et al. (1982) recruited participants who self-reported either extensive, or very little experience with athletic activities, and then trained participants on a coincident timing task under with either a single constant training velocity, with one of several varied training procedures. Unsurprisingly, athlete participants had superior performance during training, regardless of condition, and training performance was superior for all subjects in the constant group. Of greater interest is the pattern of testing results from novel transfer conditions. Among the athlete-participants, transfer performance was best for those who received variable training. Non-athletes showed the opposite pattern, with superior performance for those who had constant training.

### The current work

The overarching purpose of this dissertation is to investigate the effects of training variability on learning and generalization within visuomotor skill learning and function learning. Our investigations is structured into two main projects, each employing distinct experimental paradigms and computational modeling frameworks to elucidate how and when variability in training enhances or impedes subsequent generalization.

In Project 1, we investigated the influence of varied practice in a simple visuomotor projectile launching task. Experiments 1 and 2 compared the performance of constant and varied training groups to assess potential benefits of variability on transfer to novel testing conditions. To account for the observed empirical effects, we introduced the Instance-based Generalization with Adaptive Similarity (IGAS) model. IGAS provides a novel computational approach for quantifying the similarity between training experiences and transfer conditions, while also allowing for variability to influence the generalization gradient itself.

Project 2 shifted focus to the domain of function learning by employing a visuomotor extrapolation task. Across three experiments, we examined how constant and varied training regimes affected learning, discrimination between stimuli, and the ability to extrapolate to novel regions of the function’s input space. To model human performance in this task, we fit the influential Associative Learning Model (ALM) and the Extrapolation-Association Model (EXAM) to individual participant data using advanced Bayesian parameter estimation techniques.

# Project 1

# Abstract

Exposing learners to variability during training has been demonstrated to improve performance in subsequent transfer testing. Such variability benefits are often accounted for by assuming that learners are developing some general task schema or structure. However much of this research has neglected to account for differences in similarity between varied and constant training conditions. In a between-groups manipulation, we trained participants on a simple projectile launching task, with either varied or constant conditions. We replicate previous findings showing a transfer advantage of varied over constant training. Furthermore, we show that a standard similarity model is insufficient to account for the benefits of variation, but, if the model is adjusted to assume that varied learners are tuned towards a broader generalization gradient, then a similarity-based model is sufficient to explain the observed benefits of variation. Our results therefore suggest that some variability benefits can be accommodated within instance-based models without positing the learning of some schemata or structure.

## Similarity and instance-based approaches to transfer of learning

Early models of learning often assumed that discrete experiences with some task or category were not stored individually in memory, but instead promoted the formation of a summary representation, often referred to as a prototype or schema, and that exposure to novel examples would then prompt the retrieval of whichever preexisting prototype was most similar. In addition to being a landmark study on the influence of training variability, Posner & Keele (1968) (described above) also put forward an influential argument concerning the nature of the mental representations acquired during learning - namely that learners tend to abstract a prototype, or aggregate representation of the dot pattern categories, rather than encoding each individual stimuli. Recall that participants are trained on only on distortions of the category prototypes (e.g. low, medium or high distortions), never encountering the exact prototypes during the training stage. Then, in the testing phase, participants are tested with the prototype patterns, their old training items, and novel low, medium and high distortions. The authors found that participants had the highest testing accuracy for the previously unseen prototype patterns, followed by the old training items, and then the novel low, medium and high distortions. The authors interpreted this pattern as evidence that participants had acquired prototype representation of the category, as opposed to storing each individual training instance, and that generalization was based on the similarity of the testing items to the learned prototype representations. Posner & Keele (1968) has been extremely influential, and continues to be cited in contemporary research as clear evidence that prototype abstraction underlies the benefits of varied training. It’s also referenced as a key influence in the development of “Schema Theory of Motor Learning” Schmidt (1975), which in turn influenced decades of research on the potential benefits of varied training in motor skill learning. However a number of the core assumptions utilized by Posner & Keele (1968) were later called into question both empirically and with competing theoretical accounts (Hintzman, 1984, 1986; Knapp & Anderson, 1984; McClelland & Rumelhart, 1985; Nosofsky & Kruschke, 1992; Palmeri & Nosofsky, 2001; Zaki & Nosofsky, 2007). Palmeri & Nosofsky (2001) demonstrated the both the dangers of assuming that psychological representations mimic the metric stimulus space, as well the viability of models with simpler representational assumptions. These authors conducted a near replication of the Posner & Keele (1968) study, but also had participants provide similarity judgements of the dot pattern stimuli after completing the training phase. A multidimensional scaling analysis of the similarity judgements revelead that the psychological representations of the prototype stimuli were not located in the middle of the training stimuli, but were instead extreme points in the psychological space. The authors also demonstrated the generalization patterns of Posner & Keele (1968) could be accounted for by an exemplar-based model, without any need to assume the abstraction of a prototype.

Instance-based, or exemplar-based models generally assume that learners encode each experience with a task as a separate instance/exemplar/trace, and that each encoded trace is in turn compared against novel stimuli (Estes, 1994; Hintzman, 1984; Jamieson et al., 2022; Medin & Schaffer, 1978; Nosofsky, 1986). As the number of stored instances increases, so does the likelihood that some previously stored instance will be retrieved to aid in the performance of a novel task. Stored instances are retrieved in the context of novel stimuli or tasks if they are sufficiently similar, thus suggesting that the process of computing similarity is of central importance to generalization.

Similarity, defined in this literature as a function of psychological distance between instances or categories, has provided a successful account of generalization across numerous tasks and domains. In an influential study demonstrating an ordinal similarity effect, experimenters employed a numerosity judgment task in which participants quickly report the number of dots flashed on a screen. Performance (in terms of response times to new patterns) on novel dot configurations varied as an inverse function of their similarity to previously trained dot configurations Palmeri (1997). That is, performance was better on novel configurations moderately similar to trained configurations than to configurations with low-similarity, and also better on low-similarity configurations than to even less similar, unrelated configurations. Instance-based similarity approaches have had some success accounting for performance in certain sub-domains of motor learning (R. G. Cohen & Rosenbaum, 2004; Crump & Logan, 2010; Meigh et al., 2018; Poldrack et al., 1999; Wifall et al., 2017). Crump & Logan (2010) trained participants to type words on an unfamiliar keyboard, while constraining the letters composing the training words to a pre-specified letter set. Following training, typing speed was tested on previously experienced words composed of previously experienced letters; novel words composed of letters from the trained letter set; and novel words composed of letters from an untrained letter set. Consistent with an instance-based account, transfer performance was graded such that participants were fastest at typing the words they had previously trained on, followed by novel words composed of letters they had trained on, and slowest performance for new words composed of untrained letters.

## Issues with Previous Research

Although the benefits of training variation in visuomotor skill learning have been observed many times, null findings have also been repeatedly found, leading some researchers to question the veracity of the variability of practice hypothesis (Newell, 2003; Van Rossum, 1990). Critics have also pointed out that investigations of the effects of training variability, of the sort described above, often fail to control for the effect of similarity between training and testing conditions. For training tasks in which participants have numerous degrees of freedom (e.g. projectile throwing tasks where participants control the x and y velocity of the projectile), varied groups are likely to experience a wider range of the task space over the course of their training (e.g. more unique combinations of x and y velocities). Experimenters may attempt to account for this possibility by ensuring that the training location(s) of the varied and constant groups are an equal distance away from the eventual transfer locations, such that their training throws are, on average, equally similar to throws that would lead to good performance at the transfer locations. However, even this level of experimental control may still be insufficient to rule out the effect of similarity on transfer. Given that psychological similarity is typically best described as either a Gaussian or exponentially decaying function of psychological distance (Ennis et al., 1988; Ghahramani et al., 1996; Logan, 1988; Nosofsky, 1992; Shepard, 1987; Thoroughman & Taylor, 2005), it is plausible that a subset of the most similar training instances could have a disproportionate impact on generalization to transfer conditions, even if the average distance between training and transfer conditions is identical between groups. **?@fig-toy-model1** demonstrates the consequences of a generalization gradient that drops off

# Project 2

## Introduction

A longstanding issue across both science and instruction has been to understand how various aspects of an educational curriculum or training program influence learning acquisition and generalization. One such aspect, which has received a great deal of research attention, is the variability of examples experienced during training (Raviv et al., 2022). The influence of training variation has been studied in numerous domains, including category learning (A. L. Cohen et al., 2001; Posner & Keele, 1968), visuomotor learning (Berniker et al., 2014; Schmidt, 1975), language learning (Perry et al., 2010), and education (Braithwaite & Goldstone, 2015; Guo et al., 2014). The pattern of results is complex, with numerous studies finding both beneficial (Braun et al., 2009; Catalano & Kleiner, 1984; Roller et al., 2001), as well as null or negative effects (Brekelmans et al., 2022; Hu & Nosofsky, 2024; Van Rossum, 1990). The present study seeks to contribute to the large body of existing research by examining the influence of variability in visuomotor function learning - a domain in which it has been relatively under-studied.

### Function Learning and Extrapolation

The study of human function learning investigates how people learn relationships between continuous input and output values. Function learning is studied both in tasks where individuals are exposed to a sequence of input/output pairs (DeLosh et al., 1997; McDaniel et al., 2013), or situations where observers are presented with an incomplete scatterplot or line graph and make predictions about regions of the plot that don’t contain data (Ciccione & Dehaene, 2021; Courrieu, 2012; Said & Fischer, 2021; Schulz et al., 2020).

Carroll (1963) conducted the earliest work on function learning. Input stimuli and output responses were both lines of varying length. The correct output response was related to the length of the input line by a linear, quadratic, or random function. Participants in the linear and quadratic performed above chance levels during extrapolation testing, with those in the linear condition performing the best overall. Carroll argued that these results were best explained by a ruled based model wherein learners form an abstract representation of the underlying function. Subsequent work by Brehmer (1974),testing a wider array of functional forms, provided further evidence for superior extrapolation in tasks with linear functions. Brehmer argued that individuals start out with an assumption of a linear function, but given sufficient error will progressively test alternative hypothesis with polynomials of greater degree. Koh & Meyer (1991) employed a visuomotor function learning task, wherein participants were trained on examples from an unknown function relating the length of an input line to the duration of a response (time between keystrokes). In this domain, participants performed best when the relation between line length and response duration was determined by a power, as opposed to linear function. Koh & Meyer developed the log-polynomial adaptive-regression model to account for their results.

The first significant challenge to the rule-based accounts of function learning was put forth by DeLosh et al. (1997) . In their task, participants learned to associate stimulus magnitudes with response magnitudes that were related via either linear, exponential, or quadratic function. Participants approached ceiling performance by the end of training in each function condition, and were able to correctly respond in interpolation testing trials. All three conditions demonstrated some capacity for extrapolation, however participants in the linear condition tended to underestimate the true function, while exponential and quadratic participants reliably overestimated the true function on extrapolation trials. Extrapolation and interpolation performance are depicted in **?@fig-delosh-extrap**.

The authors evaluated both of the rule-based models introduced in earlier research (with some modifications enabling trial-by-trial learning). The polynomial hypothesis testing model (Brehmer, 1974; Carroll, 1963) tended to mimic the true function closely in extrapolation, and thus offered a poor account of the human data. The log-polynomial adaptive regression model (Koh & Meyer, 1991) was able to mimic some of the systematic deviations produced by human subjects, but also predicted overestimation in cases where underestimation occurred.

The authors also introduced two new function-learning models. The Associative Learning Model (ALM) and the extrapolation-association model (EXAM). ALM is a two layer connectionist model adapted from the ALCOVE model in the category learning literature (Kruschke, 1992). ALM belongs to the general class of radial-basis function neural networks, and can be considered a similarity-based model in the sense that the nodes in the input layer of the network are activated as a function of distance. The EXAM model retains the same similarity based activation and associative learning mechanisms as ALM, while being augmented with a linear rule response mechanism. When presented with novel stimuli, EXAM will retrieve the most similar input-output examples encountered during training, and from those examples compute a local slope. ALM was able to provide a good account of participant training and interpolation data in all three function conditions, however it was unable to extrapolate. EXAM, on the other hand, was able to reproduce both the extrapolation underestimation, as well as the quadratic and exponential overestimation patterns exhibited by the human participants. Subsequent research identified some limitations in EXAM’s ability to account for cases where human participants learn and extrapolate sinusoidal function Bott & Heit (2004) or to scenarios where different functions apply to different regions of the input space Kalish et al. (2004), though EXAM has been shown to provide a good account of human learning and extrapolation in tasks with bi-linear, V shaped input spaces Mcdaniel et al. (2009).

### Variability and Function Learning

The influence of variability on function learning tasks has received relatively little attention. The study by DeLosh et al. (1997) (described in detail above) did include a variability manipulation (referred to as density in their paper), wherein participants were trained with either either 8, 20, or 50 unique input-output pairs, with the total number of training trials held constant. They found a minimal influence of variability on training performance, and no difference between groups in interpolation or extrapolation, with all three variability conditions displaying accurate interpolation, and linearly biased extrapolation that was well accounted for by the EXAM model.

### Overview Of Present Study

The present study investigates the influence of training variability on learning, generalization, and extrapolation in a uni-dimensional visuomotor function learning task. To the best of our knowledge, this research is the first to employ the classic constant vs. varied training manipulation, commonly used in the literature on the benefits of variability, in the context of a uni-dimensional function learning task. Across three experiments, we compare constant and varied training conditions in terms of learning performance, extrapolation accuracy, and the ability to reliably discriminate between stimuli.

To account for the empirical results, we will apply a series of computational models, including the Associative Learning Model (ALM) and the Extrapolation-Association Model (EXAM). Notably, this study is the first to employ approximate Bayesian computation (ABC) to fit these models to individual subject data, enabling us to thoroughly investigate the full range of posterior predictions of each model, and to examine the ability of these influential models of function learning to account for both the group level and individual level data.

## Experiment 1

### Methods

*Participants* A total of 156 participants were recruited from the Indiana University Introductory Psychology Course. Participants were randomly assigned to one of two training conditions: varied training or constant training.

*Task.* The “Hit The Wall” (HTW) visuomotor extrapolation task task was programmed in Javascript, making heavy use of the [phaser.io](https://phaser.io/) game library. The HTW task involved launching a projectile such that it would strike the “wall” at target speed indicated at the top of the screen (see [Figure 1](#fig-htw-task)). The target velocities were given as a range, or band, of acceptable velocity values (e.g., band 800-1000). During the training stage, participants received feedback indicating whether they had hit the wall within the target velocity band, or how many units their throw was above or below from the target band. Participants were instructed that only the x velocity component of the ball was relevant to the task. The y velocity, or the location at which the ball struck the wall, had no influence on the task feedback.

|  |
| --- |
| Figure 1: The Hit the wall task. Participants launch the blue ball to hit the red wall at the target velocity band indicated at the top of the screen. The ball must be released from within the orange square - but the location of release, and the location at which the ball strikes the wall are both irrelevant to the task feedback. |

*Procedure.* All participants completed the task online. Participants were provided with a description of the experiment and indicated informed consent. [Figure 2](#fig-design-e1) illustrates the general procedure. Participants completed a total of 90 trials during the training stage. In the varied training condition, participants encountered three velocity bands (800-1000, 1000-1200, and 1200-1400). Participants in the constant training condition trained on only one velocity band (800-1000) - the closest band to what would be the novel extrapolation bands in the testing stage.

Following the training stage, participants proceeded immediately to the testing stage. Participants were tested from all six velocity bands, in two separate stages. In the novel extrapolation testing stage, participants completed “no-feedback” testing from three novel extrapolation bands (100-300, 350-550, and 600-800), with each band consisting of 15 trials. Participants were also tested from the three velocity bands that were trained by the varied condition (800-1000, 1000-1200, and 1200-1400). In the constant training condition, two of these bands were novel, while in the varied training condition, all three bands were encountered during training. The order in which participants completed the novel-extrapolation and testing-from-3-varied bands was counterbalanced across participants. A final training stage presented participants with “feedback” testing for each of the three extrapolation bands (100-300, 350-550, and 600-800).

|  |
| --- |
| Figure 2: Experiment 1 Design. Constant and Varied participants complete different training conditions. |

### Analyses Strategy

All data processing and statistical analyses were performed in R version 4.32 (Team, 2020). To assess differences between groups, we used Bayesian Mixed Effects Regression. Model fitting was performed with the brms package in R (Bürkner, 2017), and descriptive stats and tables were extracted with the BayestestR package (Makowski et al., 2019). Mixed effects regression enables us to take advantage of partial pooling, simultaneously estimating parameters at the individual and group level. Our use of Bayesian, rather than frequentist methods allows us to directly quantify the uncertainty in our parameter estimates, as well as avoiding convergence issues common to the frequentist analogues of our mixed models.

Each model was set to run with 4 chains, 5000 iterations per chain, with the first 2500 discarded as warmup chains. Rhat values were within an acceptable range, with values <=1.02 (see appendix for diagnostic plots). We used uninformative priors for the fixed effects of the model (condition and velocity band), and weakly informative Student T distributions for for the random effects. For each model, we report 1) the mean values of the posterior distribution for the parameters of interest, 2) the lower and upper credible intervals (CrI), and the probability of direction value (pd).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: **Statistical Model Specifications**. The specifications for the Bayesian regression models used in the analyses of each of the 3 experiments. Comparisons of accuracy use abosulte deviation as the dependent variable, while comparisons of discrimination use the raw velocities produced by participants as the dependent variable.   | Group Comparison | Code | Data | | --- | --- | --- | | End of Training Accuracy | brm(Abs. Deviation ~ condit) | Final Training Block | | Test Accuracy | brm(Abs. Deviation ~ condit \* bandType + (1|id) + (1|bandInt) | All Testing trials | | Band Discrimination | brm(vx ~ condit \* band +(1 + bandInt|id) | All Testing Trials | |

In each experiment we compare varied and constant conditions in terms of 1) accuracy in the final training block; 2) testing accuracy as a function of band type (trained vs. extrapolation bands); 3) extent of discrimination between all six testing bands. We quantified accuracy as the absolute deviation between the response velocity and the nearest boundary of the target band. Thus, when the target band was velocity 600-800, throws of 400, 650, and 900 would result in deviation values of 200, 0, and 100, respectively. The degree of discrimination between bands was index by fitting a linear model predicting the response velocity as a function of the target velocity. Participants who reliably discriminated between velocity bands tended to haves slope values ~1, while participants who made throws irrespective of the current target band would have slopes ~0.

### Results

|  |
| --- |
| Figure 3: Experiment 1 Training Stage. Deviations from target band across training blocks. Lower values represent greater accuracy. |

### Modelling Results

[1] "docx"

[1] FALSE

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Models errors predicting empirical data from Experiment 1 - aggregated over the full posterior distribution for each participant. Note that Fit Method refers to the subset of the data that the model was trained on, while Task Stage refers to the subset of the data that the model was evaluated on.   | Task Stage | Fit Method | ALM\_Constant | ALM\_Varied | EXAM\_Constant | EXAM\_Varied | | --- | --- | --- | --- | --- | --- | | Test | Fit to Test Data | 200 | 103 | 104 | 86 | | Test | Fit to Test & Training Data | 217 | 170 | 128 | 145 | | Test | Fit to Training Data | 468 | 291 | 273 | 298 | | Train | Fit to Test Data | 298 | 2016 | 54 | 184 | | Train | Fit to Test & Training Data | 57 | 132 | 43 | 128 | | Train | Fit to Training Data | 52 | 103 | 51 | 107 | |

### table 2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Models errors predicting empirical data - aggregated over all participants, posterior parameter values, and velocity bands. Note that Fit Method refers to the subset of the data that the model was trained on, while Task Stage refers to the subset of the data that the model was evaluated on.   | Fit\_Method | Task Stage | E2\_ALM\_Constant | E2\_ALM\_Varied | E2\_EXAM\_Constant | E2\_EXAM\_Varied | E3\_ALM\_Constant | E3\_ALM\_Varied | E3\_EXAM\_Constant | E3\_EXAM\_Varied | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Fit to Test Data | Test | 240 | 130 | 100 | 88 | 170 | 106 | 92 | 73 | | Fit to Test Data | Train | 53 | 527 | 108 | 169 | 71 | 544 | 158 | 213 | | Fit to Test & Training Data | Test | 266 | 208 | 125 | 126 | 198 | 190 | 130 | 128 | | Fit to Test & Training Data | Train | 40 | 35 | 30 | 24 | 49 | 86 | 49 | 78 | | Fit to Training Data | Test | 357 | 296 | 305 | 235 | 415 | 299 | 295 | 244 | | Fit to Training Data | Train | 43 | 23 | 43 | 23 | 51 | 64 | 52 | 65 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Results of Bayesian Regression models predicting model error as a function of Model (ALM vs. EXAM), Condition (Constant vs. Varied), and the interaction between Model and Condition. The values represent the estimate coefficient for each term, with 95% credible intervals in brackets. The intercept reflects the baseline of ALM and Constant. The other estimates indicate deviations from the baseline for the EXAM mode and varied condition. Lower values indicate better model fit.   | exp | Term | Estimate | 95% CrI Lower | 95% CrI Upper | pd | | --- | --- | --- | --- | --- | --- | | Exp 1 | Intercept | 176.3 | 156.9 | 195 | 1.00 | | Exp 1 | ModelEXAM | -88.4 | -104.5 | -72 | 1.00 | | Exp 1 | conditVaried | -37.5 | -60.4 | -14 | 1.00 | | Exp 1 | ModelEXAM:conditVaried | 60.4 | 36.2 | 84 | 1.00 | | Exp 2 | Intercept | 245.9 | 226.2 | 265 | 1.00 | | Exp 2 | ModelEXAM | -137.7 | -160.2 | -115 | 1.00 | | Exp 2 | conditVaried | -86.4 | -113.5 | -59 | 1.00 | | Exp 2 | ModelEXAM:conditVaried | 56.9 | 25.3 | 88 | 1.00 | | Exp 3 | Intercept | 164.8 | 140.1 | 189 | 1.00 | | Exp 3 | ModelEXAM | -65.7 | -86.0 | -46 | 1.00 | | Exp 3 | conditVaried | -40.6 | -75.9 | -3 | 0.98 | | Exp 3 | bandOrderReverse | 25.5 | -9.3 | 59 | 0.93 | | Exp 3 | ModelEXAM:conditVaried | 41.9 | 11.2 | 73 | 0.99 | | Exp 3 | ModelEXAM:bandOrderReverse | -7.3 | -34.5 | 21 | 0.70 | | Exp 3 | conditVaried:bandOrderReverse | 30.8 | -19.6 | 84 | 0.88 | | Exp 3 | ModelEXAM:conditVaried:bandOrderReverse | -60.6 | -101.8 | -19 | 1.00 | |

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. <https://doi.org/10.1037//0033-2909.128.4.612>

Berniker, M., Mirzaei, H., & Kording, K. P. (2014). The effects of training breadth on motor generalization. *Journal of Neurophysiology*, *112*(11), 2791–2798. <https://doi.org/10.1152/jn.00615.2013>

Bott, L., & Heit, E. (2004). Nonmonotonic Extrapolation in Function Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 38–50. <https://doi.org/10.1037/0278-7393.30.1.38>

Braithwaite, D. W., & Goldstone, R. L. (2015). Effects of Variation and Prior Knowledge on Abstract Concept Learning. *Cognition and Instruction*, *33*(3), 226–256. <https://doi.org/10.1080/07370008.2015.1067215>

Braun, D. A., Aertsen, A., Wolpert, D. M., & Mehring, C. (2009). Motor Task Variation Induces Structural Learning. *Current Biology*, *19*(4), 352–357. <https://doi.org/10.1016/j.cub.2009.01.036>

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance*, *11*(1), 1–27. <https://doi.org/10.1016/0030-5073(74)90002-6>

Brekelmans, G., Lavan, N., Saito, H., Clayards, M., & Wonnacott, E. (2022). Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication. *Journal of Memory and Language*, *126*, 104352. <https://doi.org/10.1016/j.jml.2022.104352>

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. <https://doi.org/10.18637/jss.v080.i01>

Carroll, J. D. (1963). Functional Learning: The Learning of Continuous Functional Mappings Relating Stimulus and Response Continua. *ETS Research Bulletin Series*, *1963*(2), i–144. <https://doi.org/10.1002/j.2333-8504.1963.tb00958.x>

Catalano, J. F., & Kleiner, B. M. (1984). Distant Transfer in Coincident Timing as a Function of Variability of Practice. *Perceptual and Motor Skills*, *58*(3), 851–856. <https://doi.org/10.2466/pms.1984.58.3.851>

Ciccione, L., & Dehaene, S. (2021). Can humans perform mental regression on a graph? Accuracy and bias in the perception of scatterplots. *Cognitive Psychology*, *128*, 101406. <https://doi.org/10.1016/j.cogpsych.2021.101406>

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175. <https://doi.org/10.3758/BF03206386>

Cohen, R. G., & Rosenbaum, D. A. (2004). Where grasps are made reveals how grasps are planned: Generation and recall of motor plans. *Experimental Brain Research*, *157*(4). <https://doi.org/10.1007/s00221-004-1862-9>

Courrieu, P. (2012). Quick approximation of bivariate functions. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 89–121. <https://doi.org/10.1111/j.2044-8317.2011.02016.x>

Crump, M. J. C., & Logan, G. D. (2010). Episodic contributions to sequential control: Learning from a typist’s touch. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 662–672. <https://doi.org/10.1037/a0018390>

Del Rey, P., Wughalter, E. H., & Whitehurst, M. (1982). The Effects of Contextual Interference on Females With Varied Experience in Open Sport Skills. *Research Quarterly for Exercise and Sport*, *53*(2), 108–115. <https://doi.org/10.1080/02701367.1982.10605236>

DeLosh, E. L., McDaniel, M. A., & Busemeyer, J. R. (1997). Extrapolation: The Sine Qua Non for Abstraction in Function Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 19. <https://doi.org/10.1037/0278-7393.23.4.968>

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Ablex Publishing.

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*(4), 449–465. <https://doi.org/10.1016/0022-2496(88)90023-5>

Estes, W. K. (1994). *Classification and Cognition*. Oxford University Press.

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1996). Generalization to Local Remappings of the Visuomotor Coordinate Transformation. *Journal of Neuroscience*, *16*(21), 7085–7096. <https://doi.org/10.1523/JNEUROSCI.16-21-07085.1996>

Guadagnoli, M. A., Holcomb, W. R., & Weber, T. J. (1999). The relationship between contextual interference effects and performer expertise on the learning of a putting task. *Journal of Human Movement Studies*, *37*(1), 19–36.

Guo, J.-P., Yang, L.-Y., & Ding, Y. (2014). Effects of example variability and prior knowledge in how students learn to solve equations. *European Journal of Psychology of Education*, *29*(1), 21–42. <https://www.jstor.org/stable/43551124>

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302. <https://doi.org/10.3758/BF03195318>

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96–101. <https://doi.org/10.3758/BF03202365>

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411.

Hu, M., & Nosofsky, R. M. (2024). High-variability training does not enhance generalization in the prototype-distortion paradigm. *Memory & Cognition*, 1–16. <https://doi.org/10.3758/s13421-023-01516-1>

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology*, *1*(3), 174–183. <https://doi.org/10.1038/s44159-022-00025-3>

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of Linear Experts: Knowledge Partitioning and Function Learning. *Psychological Review*, *111*(4), 1072–1099. <https://doi.org/10.1037/0033-295X.111.4.1072>

Knapp, A. K., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 616–637.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 811. <https://doi.org/10.1037/0278-7393.17.5.811>

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of Category Learning. *Psychological Review*, *99*(1). <https://doi.org/10.1037/0033-295X.99.1.22>

Lavan, N., Knight, S., Hazan, V., & McGettigan, C. (2019). The effects of high variability training on voice identity learning. *Cognition*, *193*, 104026. <https://doi.org/10.1016/j.cognition.2019.104026>

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. <https://doi.org/10.1037/0033-295X.95.4.492>

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. <https://doi.org/10.21105/joss.01541>

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

Mcdaniel, M. A., Dimperio, E., Griego, J. A., & Busemeyer, J. R. (2009). Predicting transfer performance: A comparison of competing function learning models. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 173–195. <https://doi.org/10.1037/a0013982>

McDaniel, M. A., Fadler, C. L., & Pashler, H. (2013). Effects of spaced versus massed training in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1417–1432. <https://doi.org/10.1037/a0032184>

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. *Psychological Review*, *85*(3), 207. <https://doi.org/10.1037/0033-295X.85.3.207>

Meigh, K. M., Shaiman, S., Tompkins, C. A., Abbott, K. V., & Nokes-Malach, T. (2018). What memory representation is acquired during nonword speech production learning? The influence of stimulus features and training modality on nonword encoding. *Cogent Psychology*, *5*(1), 1493714. <https://doi.org/10.1080/23311908.2018.1493714>

Newell, K. M. (2003). Schema Theory (1975): Retrospectives and Prospectives. *Research Quarterly for Exercise and Sport*, *74*(4), 383–388. <https://doi.org/10.1080/02701367.2003.10609108>

North, J. S., Bezodis, N. E., Murphy, C. P., Runswick, O. R., Pocock, C., & Roca, A. (2019). The effect of consistent and varied follow-through practice schedules on learning a table tennis backhand. *Journal of Sports Sciences*, *37*(6), 613–620. <https://doi.org/10.1080/02640414.2018.1522683>

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. <https://doi.org/10.1037/0096-3445.115.1.39>

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*(1), 25–53. <https://doi.org/10.1146/annurev.ps.43.020192.000325>

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an Exemplar-Based Connectionist Model of Category Learning. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (Vol. 28, pp. 207–250). Academic Press. <https://doi.org/10.1016/S0079-7421(08)60491-0>

Palmeri, T. J. (1997). Exemplar Similarity and the Development of Automaticity. *Journal of Experimental Psychology: Human Learning and Memory*, *23*(2), 324–354.

Palmeri, T. J., & Nosofsky, R. M. (2001). Central Tendencies, Extreme Points, and Prototype Enhancement Effects in Ill-Defined Perceptual Categorization. *The Quarterly Journal of Experimental Psychology Section A*, *54*(1), 197–235. <https://doi.org/10.1080/02724980042000084>

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning. *Psychological Science*, *21*(12), 1894–1902. <https://doi.org/10.1177/0956797610389189>

Poldrack, R. A., Selco, S. L., Field, J. E., & Cohen, N. J. (1999). The relationship between skill learning and repetition priming: Experimental and computational analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 208–235. <https://doi.org/10.1037/0278-7393.25.1.208>

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363. <https://doi.org/10.1037/h0025953>

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, S1364661322000651. <https://doi.org/10.1016/j.tics.2022.03.007>

Roller, C. A., Cohen, H. S., Kimball, K. T., & Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor plasticity. *Cognitive Brain Research*, *12*(2), 341–352. <https://doi.org/10.1016/S0926-6410(01)00077-5>

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*, 1318. <https://doi.org/10.3389/fpsyg.2014.01318>

Said, N., & Fischer, H. (2021). Extrapolation accuracy underestimates rule learning: Evidence from the function-learning paradigm. *Acta Psychologica*, *218*, 103356. <https://doi.org/10.1016/j.actpsy.2021.103356>

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

Schulz, E., Quiroga, F., & Gershman, S. J. (2020). Communicating Compositional Patterns. *Open Mind*, *4*, 25–39. <https://doi.org/10.1162/opmi_a_00032>

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. <https://doi.org/10.1126/science.3629243>

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*, *41*(04), 795–820. <https://doi.org/10.1017/S0272263119000263>

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. R: A Language and Environment for Statistical Computing.

Thoroughman, K. A., & Taylor, J. A. (2005). Rapid Reshaping of Human Motor Generalization. *Journal of Neuroscience*, *25*(39), 8948–8953. <https://doi.org/10.1523/JNEUROSCI.1771-05.2005>

Van Rossum, J. H. A. (1990). Schmidt’s schema theory: The empirical base of the variability of practice hypothesis. *Human Movement Science*, *9*(3-5), 387–435. <https://doi.org/10.1016/0167-9457(90)90010-B>

Wifall, T., Buss, A. T., Farmer, T. A., Spencer, J. P., & Hazeltine, E. (2017). Reaching into response selection: Stimulus and response similarity influence central operations. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 555–568. <https://doi.org/10.1037/xhp0000301>

Wrisberg, C. A., Winter, T. P., & Kuhlman, J. S. (1987). The Variability of Practice Hypothesis: Further Tests and Methodological Discussion. *Research Quarterly for Exercise and Sport*, *58*(4), 369–374. <https://doi.org/10.1080/02701367.1987.10608114>

Zaki, S. R., & Nosofsky, R. M. (2007). A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test. *Memory & Cognition*, *35*(8), 2088–2096. <https://doi.org/10.3758/BF03192940>

Zaman, J., Chalkia, A., Zenses, A.-K., Bilgin, A. S., Beckers, T., Vervliet, B., & Boddez, Y. (2021). Perceptual variability: Implications for learning and generalization. *Psychonomic Bulletin & Review*, *28*(1), 1–19. <https://doi.org/10.3758/s13423-020-01780-1>