

Introduction

immediate

Author note

Introduction

Introduction

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.

The study of variability

Varied training has been shown to influence learning in wide array of different tasks and domains, including categorization (Hahn et al., 2005; Maddox & Filoteo, 2011; Nosofsky et al., 2019; Posner & Keele, 1968), language learning (Jones & Brandt, 2020; Perry et al., 2010; Twomey et al., 2018; Wonnacott et al., 2012), anagram completion (Goode et al., 2008), trajectory extrapolation (Fulvio et al., 2014), task switching (Sabah et al., 2019), associative learning (Lee et al., 2019), visual search (George & Egner, 2021; Gonzalez & Madhavan, 2011; Kelley & Yantis, 2009), voice identity learning (Lavan et al., 2019), simple motor learning (Braun et al., 2009; Kerr & Booth, 1978; Roller et al., 2001; Willey & Liu, 2018a), sports training (Green et al., 1995; North et al., 2019),

and training on a complex video game (Seow et al., 2019). See Czyż (2021) or Raviv et al. (2022) for more detailed reviews.

Research on the effects of varied training typically manipulates variability in one of two ways. In the first approach, a high variability group is exposed to a greater number of unique instances during training, while a low variability group receives fewer unique instances with more repetitions. Alternatively, both groups may receive the same number of unique instances, but the high variability group's instances are more widely distributed or spread out in the relevant psychological space, while the low variability group's instances are clustered more tightly together. Researchers then compare the training groups in terms of their performance during the training phase, as well as their generalization performance during a testing phase. Researchers will usually compare the performance of the two groups during both the training phase and a subsequent testing phase. The primary theoretical interest is often to assess the influence of training variability on generalization to novel testing items or conditions. However, the test may also include some or all of the items that were used during the training stage, allowing for an assessment of whether the variability manipulation influenced the learning of the trained items themselves, or to make it easy to measure how much performance degrades as a function of how far away testing items are from the training items.

The influence of training variation has received a large amount of attention in the domain of sensorimotor skill learning. Much of this research has been influenced by the work of (Schmidt (1975)), who proposed a schema-based account of motor learning as an attempt to address the longstanding problem of how novel movements are produced. Schema theory presumes that learners possess general motor programs for a class of movements (e.g. an underhand throw). When called up for use motor programs are parameterized by schema rules which determine

how the motor program is parameterized or scaled to the particular demands of the current task. Schema theory predicts that variable training facilitates the formation of more robust schemas, which will result in improved generalization or transfer. Experiments that test this hypothesis are often designed to compare the transfer performance of a constant-trained group against that of a varied-trained group. Both groups train on the same task, but the varied group practices with multiple instances along some task-relevant dimension that remains invariant for the constant group. For example, studies using a projectile throwing task might assign participants to either constant training that practicing throwing from a single location, and a varied group that throws from multiple locations. Following training, both groups are then tested from novel throwing locations (Pacheco & Newell, 2018; Pigott & Shapiro, 1984; Willey & Liu, 2018a; Wulf, 1991).

One of the earliest, and still often cited investigations of Schmidt's benefits of variability hypothesis was the work of Kerr and Booth (1978). Two groups of children, aged 8 and 12, were assigned to either constant or varied training of a bean bag throwing task. The constant group practiced throwing a bean-bag at a small target placed 3 feet in front of them, and the varied group practiced throwing from a distance of both 2 feet and 4 feet. Participants were blindfolded and unable to see the target while making each throw but would receive feedback by looking at where the beanbag had landed in between each training trial. 12 weeks later, all of the children were given a final test from a distance of 3 feet which was novel for the varied participants and repeated for the constant participants. Participants were also blindfolded for testing and did not receive trial by trial feedback in this stage. In both age groups, participants performed significantly better in the varied condition than the constant condition, though the effect was larger for the younger, 8-year-old children. This result offers a particularly compelling example of the merits of varied practice, given that the varied group was able to outperform the constant group even from

the home turf location where one may have expected the constant group to have the strongest advantage. A similar pattern of results was observed in another study wherein varied participants trained with tennis, squash, badminton, and short-tennis rackets were compared against constant subjects trained with only a tennis racket (Green et al., 1995). One of the testing conditions had subjects repeat the use of the tennis racket, which had been used on all 128 training trials for the constant group, and only 32 training trials for the varied group. Nevertheless, the varied group outperformed the constant group when using the tennis racket at testing, and also performed better in conditions with several novel racket lengths. However, as is the case with many of the patterns commonly observed in the “benefits of variability” literature, the pattern wherein the varied group outperforms the constant group even from the constant group’s home turf has not been consistently replicated. One recent study attempted a near replication of the Kerr & Booth study (Willey & Liu, 2018b), having subjects throw beanbags at a target, with the varied group training from positions (5 and 9 feet) on either side of the constant group (7 feet). This study did not find a varied advantage from the constant training position, though the varied group did perform better at distances novel to both groups. However, this study diverged from the original in that the participants were adults; and the amount of training was much greater (20 sessions with 60 practice trials each, spread out over 5-7 weeks).

Pitting varied against constant practice against each other on the home turf of the constant group provides a compelling argument for the benefits of varied training, as well as an interesting challenge for theoretical accounts that posit generalization to occur as some function of distance. However, despite its appeal this particular contrast is relatively uncommon in the literature. It is unclear whether this may be cause for concern over publication bias, or just researchers feeling the design is too risky. A far more common design is to have separate constant groups that

each train exclusively from each of the conditions that the varied group encounters (Catalano & Kleiner, 1984; Chua et al., 2019; McCracken & Stelmach, 1977; Moxley, 1979; Newell & Shapiro, 1976), or for a single constant group to train from just one of the conditions experienced by the varied participants (Pigott & Shapiro, 1984; Roller et al., 2001; Wrisberg & McLean, 1984; Wrisberg & Mead, 1983). A less common contrast places the constant group training in a region of the task space outside of the range of examples experienced by the varied group, but distinct from the transfer condition (Wrisberg et al., 1987; Wulf & Schmidt, 1997). Of particular relevance to the current work is the early study of Catalano and Kleiner (1984), as theirs was one of the earliest studies to investigate the influence of varied vs. constant training on multiple testing locations of graded distance from the training condition. Participants were trained on coincident timing task, in which subjects observe a series of lightbulbs turning on sequentially at a consistent rate and attempt to time a button response with the onset of the final bulb. The constant groups trained with a single velocity of either 5, 7, 9, or 11 mph, while the varied group trained from all 4 of these velocities. Participants were then assigned to one of four possible generalization conditions, all of which fell outside of the range of the varied training conditions – 1, 3, 13 or 15 mph. As is often the case, the varied group performed worse during the training phase. In the testing phase, the general pattern was for all participants to perform worse as the testing conditions became further away from the training conditions, but since the drop off in performance as a function of distance was far less steep for the varied group, the authors suggested that varied training induced a decremented generalization gradient, such that the varied participants were less affected by the change between training and testing conditions.

Benefits of varied training have also been observed in many studies outside of the sensorimotor domain. Goode et al. (2008) trained participants to solve anagrams of 40 different words

ranging in length from 5 to 11 letters, with an anagram of each word repeated 3 times throughout training, for a total of 120 training trials. Although subjects in all conditions were exposed to the same 40 unique words (i.e. the solution to an anagram), participants in the varied group saw 3 different arrangements for each solution-word, such as DOLOF, FOLOD, and OOFLO for the solution word FLOOD, whereas constant subjects would train on three repetitions of LDOOF (spread evenly across training). Two different constant groups were used. Both constant groups trained with three repetitions of the same word scramble, but for constant group A, the testing phase consisted of the identical letter arrangement to that seen during training (e.g. LDOOF), whereas for constant group B, the testing phase consisted of a arrangement they had not seen during training, thus presenting them with a testing situation similar situation to the varied group. At the testing stage, the varied group outperformed both constant groups, a particularly impressive result, given that constant group A had 3 prior exposures to the word arrangement (i.e. the particular permutation of letters) which the varied group had not explicitly seen. However varied subjects in this study did not exhibit the typical decrement in the training phase typical of other varied manipulations in the literature, and actually achieved higher levels of anagram solving accuracy by the end of training than either of the constant groups – solving 2 more anagrams on average than the constant group. This might suggest that for tasks of this nature where the learner can simply get stuck with a particular word scramble, repeated exposure to the identical scramble might be less helpful towards finding the solution than being given a different arrangement of the same letters. This contention is supported by the fact that constant group A, who was tested on the identical arrangement as they experienced during training, performed no better at testing than did constant group B, who had trained on a different arrangement of the same word solution – further suggesting that there may not have been a strong identity advantage in this task.

In the domain of category learning, the constant vs. varied comparison is much less suitable. Instead, researchers will typically employ designs where all training groups encounter numerous stimuli, but one group experiences a greater number of unique exemplars (Brunstein & Gonzalez, 2011; Doyle & Hourihan, 2016; Hosch et al., 2023; Nosofsky et al., 2019; Wahlheim et al., 2012), or designs where the number of unique training exemplars is held constant, but one group trains with items that are more dispersed, or spread out across the category space (Bowman & Zeithamova, 2020; Homa & Vosburgh, 1976; Hu & Nosofsky, 2024; Maddox & Filoteo, 2011; Posner & Keele, 1968).

Much of the earlier work in this sub-area trained subjects on artificial categories, such as dot patterns (Homa & Vosburgh, 1976; Posner & Keele, 1968). A seminal study by Posner and Keele (1968) trained participants to categorize artificial dot patterns, manipulating whether learners were trained with examples clustered close to the category prototypes (i.e. low variability), or spread further away from the prototype (i.e. high-distortion, or high variability-training). Participants that received training on more highly-distorted items showed superior generalization to novel high distortion patterns in the subsequent testing phase. It should be noted that unlike the sensorimotor studies discussed earlier, the Posner and Keele (1968) study did not present low-varied and high-varied participants with an equal number of training trials, but instead had participants remain in the training stage of the experiment until they reached a criterion level of performance. This train-to-criterion procedure led to the high-variability condition participants tending to complete a larger number of training trials before switching to the testing stage. More recent work (Hu & Nosofsky, 2024), also used dot pattern categories, but matched the number of training trials across conditions. This study found the opposite pattern of results, with the

low-variability group showing superior generalization to novel high-distortion patterns (as well as novel patterns of other distortion levels).

Recent studies have also begun utilizing more complex or realistic stimuli when assessing the influence of variability on category learning. Wahlheim et al. (2012) conducted one such study. In a within-participants design, participants were trained on bird categories with either high repetitions of a few exemplars, or few repetitions of many exemplars. Across four different experiments, which were conducted to address an unrelated question on metacognitive judgements, the researchers consistently found that participants generalized better to novel species following training with more unique exemplars (i.e. higher variability), while high repetition training produced significantly better performance categorizing the specific species they had trained on. A variability advantage was also found in the relatively complex domain of rock categorization (Nosofsky et al., 2019). For 10 different rock categories, participants were trained with either many repetitions of 3 unique examples of each category, or few repetitions of 9 unique examples, with an equal number of total training trials in each group (the design also included 2 other conditions less amenable to considering the impact of variation). The high-variability group, trained with 9 unique examples, showed significantly better generalization performance than the other conditions. Moreover, the pattern of results in this study could be nicely accounted for by an extended version of the Generalized Context Model.

A distinct sub-literature within the category learning domain has examined how the variability or dispersion of the categories themselves influences generalization to ambiguous regions of the category space (e.g. the region between the two categories). The general approach is to train participants with examples from a high variability category and a low variability category. Participants are then tested with novel items located within ambiguous regions of the category space

which allow the experimenters to assess whether the difference in category variability influenced how far participants generalize the category boundaries. Cohen et al. (2001) trained subjects on two categories, one with much more variability than the other. In experiment 1, a low variability category composed of 1 instance was compared against a high-variability category of 2 instances in one condition, and 7 instances in another. In experiment 2 both categories were composed of 3 instances, but for the low-variability group the instances were clustered close to each other, whereas the high-variability groups instances were spread much further apart. Participants were tested on an ambiguous novel instance that was located in between the two trained categories. Both experiments provided evidence that participants were much more likely to categorize the novel middle stimulus into a category with greater variation.

Further observations consonant with the results of Cohen et al. (2001) have since been observed in numerous investigations (Hahn et al., 2005; Hsu & Griffiths, 2010; Perlman et al., 2012; Sakamoto et al., 2008; but see Stewart & Chater, 2002). The results of Sakamoto et al. (2008) are noteworthy. They first reproduced the basic finding of participants being more likely to categorize an unknown middle stimulus into a training category with higher variability. In a second experiment, they held the variability between the two training categories constant and instead manipulated the training sequence, such that the examples of one category appeared in an ordered fashion, with very small changes from one example to the other (the stimuli were lines that varied only in length), whereas examples in the alternate category were shown in a random order and thus included larger jumps in the stimulus space from trial to trial. They found that the middle stimulus was more likely to be categorized into the category that had been learned with a random sequence, which was attributed to an increased perception of variability which resulted from the larger trial to trial discrepancies.

The work of Hahn et al. (2005), is also of particular interest to the present work. Their experimental design was similar to previous studies, but they included a larger set of testing items which were used to assess generalization both between the two training categories as well as novel items located in the outer edges of the training categories. During generalization testing, participants were given the option to respond with “neither”, in addition to responses to the two training categories. The “neither” response was included to test how far away in the stimulus space participants would continue to categorize novel items as belonging to a trained category. Consistent with prior findings, high-variability training resulted in an increased probability of categorizing items in between the training categories as belong to the high variability category. Additionally, participants trained with higher variability also extended the category boundary further out into the periphery than participants trained with a lower variability category were willing to do. The author compared a variety of similarity-based models based around the Generalized Context Model (Nosofsky, 1986) to account for their results, manipulating whether a response-bias or similarity-scaling parameter was fit separately between variability conditions. No improvement in model fit was found by allowing the response-bias parameter to differ between groups, however the model performance did improvement significantly when the similarity scaling parameter was fit separately. The best fitting similarity-scaling parameters were such that the high-variability group was less sensitive to the distances between stimuli, resulting in greater similarity values between their training items and testing items. This model accounted for both the extended generalization gradients of the varied participants, and also for their poorer performance in a recognition condition.

Variability has also been examined in the learning of higher-order linguistic categories (Perry et al., 2010). In nine training sessions spread out over nine weeks infants were trained on

object labels in a naturalistic play setting. All infants were introduced to three novel objects of the same category, with participants in the “tight” condition being exposed to three similar exemplars of the category, and participants in the varied condition being exposed to three dissimilar objects of the same category. Importantly, the similarity of the objects was carefully controlled for by having a separate group of adult subjects provide pairwise similarity judgements of the category objects prior to the study onset. Multidimensional scaling was then performed to obtain the coordinates of the objects psychological space, and out of the 10 objects for each category, the 3 most similar objects were selected for the tight group and the three least similar objects for the varied group, with the leftover four objects being retained for testing. By the end of the nine weeks, all of the infants had learned the labels of the training objects. In the testing phase, the varied group demonstrated superior ability to correctly generalize the object labels to untrained exemplars of the same category. More interesting was the superior performance of the varied group on a higher order generalization task – such that they were able to appropriately generalize the bias they had learned during training for attending to the shape of objects to novel solid objects, but not to non-solids. The tight training group, on the other hand, tended to overgeneralize the shape bias, leading the researchers to suggest that the varied training induced a more context-sensitive understanding of when to apply their knowledge.

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.

References

- 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>
- Bowman, C. R., & Zeithamova, D. (2020). Training set coherence and set size effects on concept generalization and recognition. *Journal of experimental psychology. Learning, memory, and cognition*, 46(8), 1442–1464. <https://doi.org/10.1037/xlm0000824>
- 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>
- Brunstein, A., & Gonzalez, C. (2011). Preparing for novelty with diverse training. *Applied Cognitive Psychology*, 25(5), 682–691. <https://doi.org/10.1002/acp.1739>
- 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>
- Chua, L.-K., Dimapilis, M. K., Iwatsuki, T., Abdollahipour, R., Lewthwaite, R., & Wulf, G. (2019). Practice variability promotes an external focus of attention and enhances motor skill

- learning. *Human Movement Science*, 64, 307–319. <https://doi.org/10.1016/j.humov.2019.02.015>
- 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>
- Czyż, S. H. (2021). Variability of Practice, Information Processing, and Decision Making—How Much Do We Know? *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.639131>
- 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.
- Doyle, M. E., & Hourihan, K. L. (2016). Metacognitive monitoring during category learning: How success affects future behaviour: Memory. *Memory*, 24(9), 1197–1207. <https://doi.org/10.1080/09658211.2015.1086805>
- Fulvio, J. M., Green, C. S., & Schrater, P. R. (2014). Task-Specific Response Strategy Selection on the Basis of Recent Training Experience. *PLOS Computational Biology*, 10(1), e1003425. <https://doi.org/10.1371/journal.pcbi.1003425>
- George, N., & Egner, T. (2021). Stimulus variability and task relevance modulate binding-learning. *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-021-02338-6>

- Gonzalez, C., & Madhavan, P. (2011). Diversity during training enhances detection of novel stimuli. *Journal of Cognitive Psychology*, 23(3), 342–350. <https://doi.org/10.1080/20445911.2011.507187>
- Goode, M. K., Geraci, L., & Roediger, H. L. (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review*, 15(3), 662–666. <https://doi.org/10.3758/PBR.15.3.662>
- Green, D. P., Whitehead, J., & Sugden, D. A. (1995). Practice Variability and Transfer of a Racket Skill. *Perceptual and Motor Skills*, 81(3_suppl), 1275–1281. <https://doi.org/10.2466/pms.1995.81.3f.1275>
- Guadagnoli, M. A., Holcomb, W., & Weber, T. (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.
- 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>
- Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, 2(3), 322–330. <https://doi.org/10.1037/0278-7393.2.3.322>
- Hosch, A.-K., Wirtz, P., & von Helversen, B. (2023). Prior Experience of Variability Influences Generalization of Unspecified Categories. *Quarterly Journal of Experimental Psychology* (2006), 17470218231210491. <https://doi.org/10.1177/17470218231210491>
<https://osf.io/tdjy2/>.

- Hsu, A. S., & Griffiths, T. L. (2010). Effects of generative and discriminative learning on use of category variability. *32nd annual conference of the cognitive science society*, 7.
- 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>
<https://osf.io/c6wea/>.
- Jones, S. D., & Brandt, S. (2020). Density and Distinctiveness in Early Word Learning: Evidence From Neural Network Simulations. *Cognitive Science*, 44(1), e12812. <https://doi.org/10.1111/cogs.12812>
- Kelley, T. A., & Yantis, S. (2009). Learning to attend: Effects of practice on information selection. *Journal of Vision*, 9(7), 16. <https://doi.org/10.1167/9.7.16>
- Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and motor skills*, 46(2), 395–401. <https://doi.org/10.1177/003151257804600201>
- 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>
- Lee, J. C., Lovibond, P. F., & Hayes, B. K. (2019). Evidential diversity increases generalisation in predictive learning. *Quarterly Journal of Experimental Psychology*, 72(11), 2647–2657. <https://doi.org/10.1177/1747021819857065>
- Maddox, W. T., & Filoteo, J. V. (2011). Stimulus range and discontinuity effects on information-integration category learning and generalization. *Attention, Perception, & Psychophysics*, 73(4), 1279–1295. <https://doi.org/10.3758/s13414-011-0101-2>

- McCracken, H. D., & Stelmach, G. E. (1977). A Test of the Schema Theory of Discrete Motor Learning. *Journal of Motor Behavior*, 9(3), 193–201. <https://doi.org/10.1080/00222895.1977.10735109>
- Moxley, S. (1979). Schema: The Variability of Practice Hypothesis. *Journal of Motor Behavior*, 11(1), 65–70. <https://doi.org/10.1080/00222895.1979.10735173>
- Newell, K., & Shapiro, D. C. (1976). Variability of Practice and Transfer of Training: Some Evidence Toward a Schema View of Motor Learning. *Journal of Motor Behavior*, 8(3), 233–243. <https://doi.org/10.1080/00222895.1976.10735077>
- 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., Sanders, C. A., Zhu, X., & McDaniel, M. A. (2019). Model-guided search for optimal natural-science-category training exemplars: A work in progress. *Psychonomic Bulletin & Review*, 26(1), 48–76. <https://doi.org/10.3758/s13423-018-1508-8>
- Pacheco, M. M., & Newell, K. M. (2018). Learning a specific, individual and generalizable coordination function: Evaluating the variability of practice hypothesis in motor learning. *Experimental Brain Research*, 236(12), 3307–3318. <https://doi.org/10.1007/s00221-018-5383-3>

- Perlman, A., Hahn, U., Edwards, D. J., & Pothos, E. M. (2012). Further attempts to clarify the importance of category variability for categorisation. *Journal of Cognitive Psychology*, 24(2), 203–220. <https://doi.org/10.1080/20445911.2011.613818>
- 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>
- Pigott, R. E., & Shapiro, D. C. (1984). Motor Schema: The Structure of the Variability Session. *Research Quarterly for Exercise and Sport*, 55(1), 41–45. <https://doi.org/10.1080/02701367.1984.10605353>
- 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](https://doi.org/10.1016/S0926-6410(01)00077-5)
- Sabah, K., Dolk, T., Meiran, N., & Dreisbach, G. (2019). When less is more: Costs and benefits of varied vs. fixed content and structure in short-term task switching training. *Psychological Research*, 83(7), 1531–1542. <https://doi.org/10.1007/s00426-018-1006-7>
- 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>

- Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory & Cognition*, 36(6), 1057–1065. <https://doi.org/10.3758/MC.36.6.1057>
- Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, 82(4), 225–260. <https://doi.org/10.1037/h0076770>
- Seow, R. Y. T., Betts, S., & Anderson, J. R. (2019). Transfer effects of varied practice and adaptation to changes in complex skill acquisition. *Proceedings of the 17th International Conference on Cognitive Modelling*, 222–227.
- 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>
- Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 893–907. <https://doi.org/10.1037//0278-7393.28.5.893>
- Twomey, K. E., Ma, L., & Westermann, G. (2018). All the Right Noises: Background Variability Helps Early Word Learning. *Cognitive Science*, 42(S2), 413–438. <https://doi.org/10.1111/cogs.12539>
- Van Rossum, J. H. (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](https://doi.org/10.1016/0167-9457(90)90010-B)
- Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: Evidence for variability neglect. *Memory & Cognition*, 40(5), 703–716. <https://doi.org/10.3758/s13421-011-0180-2>

- Willey, C. R., & Liu, Z. (2018a). Limited generalization with varied, as compared to specific, practice in short-term motor learning. *Acta Psychologica*, 182, 39–45. <https://doi.org/10.1016/j.actpsy.2017.11.008>
- Willey, C. R., & Liu, Z. (2018b). Long-term motor learning: Effects of varied and specific practice. *Vision Research*, 152, 10–16. <https://doi.org/10.1016/j.visres.2017.03.012>
- Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5year olds. *Journal of Memory and Language*, 66(3), 458–478. <https://doi.org/10.1016/j.jml.2011.11.004>
- Wrisberg, C. A., & McLean, E. (1984). Training for the production of novel timing movements: Contextual considerations. *Psychological Research*, 46(1-2), 169–176. <https://doi.org/10.1007/BF00308601>
- Wrisberg, C. A., & Mead, B. J. (1983). Developing Coincident Timing Skill in Children: A Comparison of Training Methods. *Research Quarterly for Exercise and Sport*, 54(1), 67–74. <https://doi.org/10.1080/02701367.1983.10605274>
- 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>
- Wulf, G. (1991). The effect of type of practice on motor learning in children. *Applied Cognitive Psychology*, 5(2), 123–134. <https://doi.org/10.1002/acp.2350050206>
- Wulf, G., & Schmidt, R. (1997). Variability of Practice and Implicit Motor Learning.
- 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>