Keywords: Generalization, Representation, Memory, Abstraction

The Generalization Artist - Or: How Can We Characterize Human Generalization?

Mirko Thalmann[1] and Eric Schulz[1]

[1]Institute for Human-Centered AI at Helmholtz Center for Computational Health

Ingolstädter Landstr. 1

85764 Neuherberg - Germany

Author Note

Correspondence concerning this article should be addressed to Mirko Thalmann, Institute for Human-Centered AI at Helmholtz Center for Computational Health, Ingolstädter Landstr. 1, 85764 Neuherberg - Germany.

E-mail: mirkothalmann@hotmail.com

The Generalization Artist - Or: How Can We Characterize Human Generalization?

A common understanding is that people are generalization artists: they require far fewer experience to generalize their knowledge when compared to contemporary AI systems, i.e. deep neural network models. Here, we summarize evidence in favor and against this notion. We propose three stages, determining how people generalize. First, people must infer what aspects of an environment are relevant for a task. Second, they need to develop a strategy to solve it. And third, while repeatedly carrying out the task, mental representations required to solve the task change. Mechanisms in all three stages can decrease the correspondence between the structure of the actual task and how people solve it. People use their lifelong experiences to constrain what features of a task are important, and they tend to start solving the task with a simple rule. On average, these decisions correspond well to natural circumstances. The true artistry of human generalization is therefore not a general ability to generalize well in any scenario, but to establish and revise efficient representations in the face of limited processing capacity.

1 The Generalization Artist

The field of artificial intelligence often looks to humans as role models, particularly for their remarkable ability to generalize. Humans are said to glean extensive insights from minimal data, easily applying their knowledge to new situations and objects with little prior experience. For example, when deep neural networks for object recognition are benchmarked against human participants, humans outperform convolutional neural networks on a variety of tasks (Geirhos et al., 2018) (e.g., adding noise to images affects humans much less). In line with this evidence is the observation of one-shot categorization (Feldman, 1992): People define categories from only one observed object and generalize accordingly, in stark contrast to learning in deep neural networks, which typically requires many more examples (e.g., Lake et al., 2017). Observations such as these have led to the portrayal of people as "generalization artists", gifted with the extraordinary ability to extrapolate from limited exposure or practice. But is this reputation deserved? How proficient are humans at generalizing knowledge to unfamiliar stimuli? To explore this, we examine evidence both supporting and challenging this view, focusing on areas such as absolute identification, category learning, function learning, reinforcement learning, language, memory, and cognitive and physical training. Ultimately, we argue that humans are indeed generalization artists—not because they always excel, but because they possess the ability to form effective representations, choose adaptive strategies, and revise their representations.

2 Evidence in Favor of the Generalization Artist

In the area of *absolute identification*, Shepard (1987) argued that the first law of psychology should be the law of generalization, since no stimulus in our environment is experienced in exactly the same circumstances, but still can be recognized as one and the same object. He showed that the probability of perceiving two stimuli as the same increases monotonically with the psychological similarity between the two stimuli. The same monotonic relationship may not hold without a transformation of the features describing the object from physical space (e.g., pitch) to psychological space (i.e., mental representation of pitch).

Moving from object generalization to *category learning*, a plethora of studies has shown that people can learn to categorize individual training stimuli into their respective category given their object properties. Importantly, stimuli not observed during training are categorized during a transfer test with an accuracy well above chance (e.g., Johansen & Palmeri, 2002; Nosofsky, 1986). People learn and generalize well when they are asked to infer categories determined by the relation between stimuli (e.g., Goldwater et al., 2018). The results look again similarly, when extending the discrete case of learning categories to *learning functions*, relating values from continuous feature dimensions (e.g., dosage of a poison) to continuous outcomes (e.g., symptom severity). Here, it has been shown that people generalize to unobserved stimuli with high accuracy in the interpolation region (i.e., values between observed values) and with performance well above chance in the extrapolation region (i.e., values outside of the observed range, e.g., DeLosh et al., 1997).

People behave similarly proficient in *reinforcement learning tasks*, in which they are instructed to collect as many rewards as possible in a limited number of choices. For example, the participants of Wu et al. (2020) learned to make use of correlations between continuous feature dimensions (e.g., tilt of a gabor patch) and rewards. Furthermore, they used their knowledge to generalize to feature values unobserved during learning. Jagadish et al. (2023) pushed this a step further; they showed that people are able to generalize to a composition of functions (i.e., adding a periodic function to a linear function) mapping response keys to rewards with remarkable accuracy on the first trial with practice only on the individual functions but without practice on the composite function.

In the area of *language* it has been suggested that abstracted knowledge helps us to understand and produce language via grammatical templates (e.g., a subject - verb - object sequence). Supporting that claim, Marcus et al. (1999) showed that already seven-month-old infants detected abstract sequential patterns (i.e., an ABA sequence, with A and B representing syllable placeholders) in a transfer sequence, which was populated with previously unobserved syllables. Abstracted sequential patterns are also helpful in the domain of *memory*: Wu et al. (2023) showed that humans extract abstracted knowledge out of patterned sequences, which improved their short-term memory performance for sequences following the same pattern but consisting of novel items. In the field of *motor skill and expert performance*, Fransen et al. (2012) showed that having practiced multiple sports as compared to a single sport improved standardized test scores of fitness and gross motor coordination in 6-12-year-olds.

To summarize, there is decisive evidence that humans generalize, be it knowledge about individual objects, categories defined by individual object features or relations between different objects, functions, or abstract patterns. Given extensive training, they sometimes even do so near optimally (e.g., Reed, 1972). While this evidence sets the expectation high for human generalization abilities, we review studies about systematic failures of generalization in the following.

3 Evidence Against the Generalization Artist

Studies in the area of *category learning* show that people over-simplify structure. For example, Vermaercke et al. (2014) trained humans and rats on a rule-based and an information-integration category structure. The latter cannot be solved with a simple one- or two-dimensional rule. Humans and rats learned both structures equally well given sufficient training. When required to categorize unobserved transfer stimuli, performance stayed roughly the same for rats in both structures. Performance, however, dropped substantially for humans in the information-integration structure, but not in the rule-based structure. In a similar information-integration category learning task, Donkin et al. (2015) showed that about a third of their participants relied on a rule-based categorization strategy, even though this strategy was clearly not the best representation of the category structure. Together, the Donkin et al. (2015) and Vermaercke et al. (2014) studies can be taken as indication that people make systematic errors when asked to infer the category of unobserved objects: they rely too heavily on rules.

Evidence from the area of *function learning* specifies this systematicity: People tend to prefer simple rules over complicated ones (see also Chater & Vitányi, 2003). For example, they learn linear functions faster than quadratic functions (Brehmer, 1974), and they simplify more complicated functions, for example by approximating a quadratic function with a linear function (DeLosh et al., 1997; Little & Shiffrin, 2009).

In the area of *cognitive training*, a large corpus of studies shows that despite large performance gains on trained tasks, performance on similar untrained tasks does not improve. For example, despite large improvements in performance in a working memory binding paradigm, performance on very similar untrained working memory tasks is unaffected (De Simoni & von Bastian, 2018). This absence of generalized training gains (e.g., Melby-Lervåg et al., 2016) has come to some surprise: Working-memory capacity and fluid intelligence correlate with .85 (Oberauer et al., 2005). It has been assumed that training gains in working memory tasks should therefore generalize to other tasks requiring reasoning or fluid intelligence. These findings cast doubt on the proposition that mastery of one task generalizes to performance on similar tasks.

Similar results come from the *motor skill and expert performance* domains. Several studies have observed that skill mastery leads to the emergence of a so-called especial skill. For example, batting accuracy of baseball players at regulation distance is substantially higher than expected from the accuracies at near-by distances (Simons et al., 2009). The latter study also demonstrated the absence of transfer/generalization to distances one foot away from regulation distance. The results from training studies contrast with the generalization artist proposition and are in line with the idea of overfitting: massive amounts of learning in a task make us just better in the trained task with exactly the used stimuli.

So, how do the diverging results about human generalization abilities, which we summarized in 1, fit together? In the following, we present a taxonomy of human generalization, which allows us to conceptualize whether we can expect generalization given some combination of training and transfer task or not.

4 A Taxonomy of Human Generalization

We propose that generalization is predominantly a function of the mental representations acquired via confrontation with a task. Representations emerge on the levels of the task and on the level of individual objects/feature dimensions. We suggest that these representations are shaped in learning episodes, which can be roughly categorized into three not mutually exclusive stages (see Figure 2). Eventually, we integrate these stages with findings from the training literatures.

4.1 Constraining The Task Representation

When people are confronted with a task, they have to constrain the dimensionality of the problem space. Even though experimenters try hard to constrain the possible task

Domain	Evidence	
	In Favor	*Against*
Absolute Identification	Psychological similarity predicts if stimuli are considered to be the same stimulus	-
Category Learning	Above chance generalization of category knowledge to unobserved stimuli	Preference for rules, even if not favorable
Function & Reinforcement Learning	Above chance generalization of function knowledge to unobserved stimuli (inter- and extrapolation)	Preference for *simple* rules (e.g., representing a quadratic function with a linear function)
Memory & Language	Use of abstracted patterns in concrete space (e.g., words, chunks) and in abstract space (e.g., nouns vs. verbs, motifs)	-
Cognitive Training	-	Absence of transfer in the face of large performance gains in trained tasks
Skilled Performance	Multi sport training favors fitness and gross motor coordination	Emergence of an especial skill (e.g., in baseball)

Figure 1

The left column presents evidence in favor of the notion of the generalization artist, the right column evidence against it.

representations via instructions, there are usually still various options how a task representation can be derived. For example, Mason et al. (2022) argue that people generate hypotheses about what aspects of a task environment are relevant. They show that slight modifications in the instructions and in the stimulus presentation affect people's task representations. This idea resonates with Feldman (1992)'s proposal that people focus on feature dimensions, which they expect to be relevant given their background knowledge, but ignore dimensions, which they assume to be irrelevant. As a consequence, individual differences in task representations arise because people differ in the features they think are task relevant.

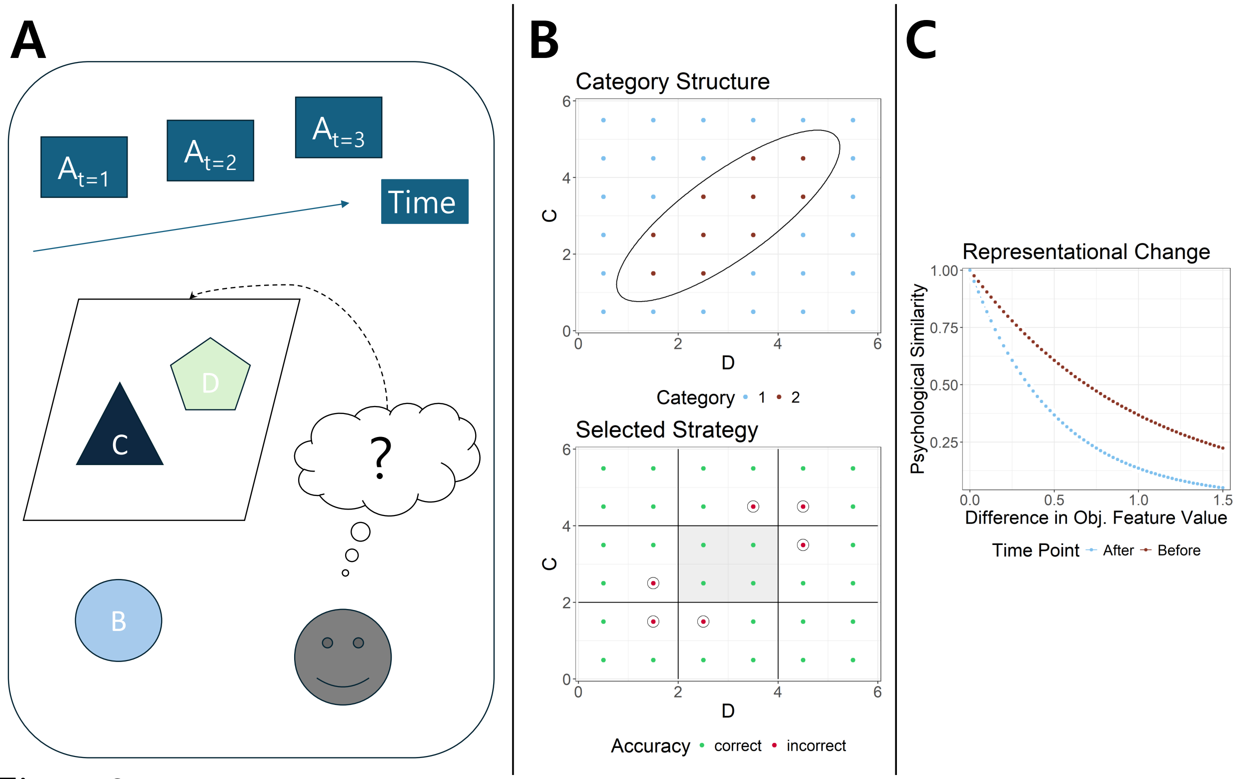

Figure 2

A: *A participant in a category learning experiment considers features A, B, C, and D to be potentially relevant in the experiment. A represents sequential information about presented stimuli, for example whether every n-th object belongs to a certain category. The person considers B to be an irrelevant feature and C and D to be likely relevant.* ***B:*** *Features C and D are indeed relevant to learn to discriminate the two categories (left panel). However, the person represents the two categories with a two-dimensional rule, which leads to systematic categorization errors (right panel).* ***C:*** *After extensive learning, representations of feature values on dimensions C and D become more precise. That is, two stimuli that differ from each other according to a fixed distance in objective feature space, are psychologically less similar after learning.*

4.2 Strategy Selection

Once people have decided upon a task representation, they must find a solution to the task. People tend to start solving a task with a rule, in particular a simple rule (Little

& Shiffrin, 2009) and they generalize according to this simple rule (Johansen & Palmeri, 2002). Simple, often verbalizable rules are cognitively little demanding and easy to learn (Feldman, 2000) and provide an effective tool to generalize to unobserved stimuli. Although a verbalizable rule may sometimes not represent the complexity of a problem well (Donkin et al., 2015), in other cases, it leads to impressive generalization performance (Nam & McClelland, 2023). From an ecological standpoint, having a simplicity preference is efficient because most everyday problems can be solved with a linear model (Jagadish et al., 2024) or a simple heuristic (Gigerenzer & Gaissmaier, 2011). We consider the empirical evidence that humans generalize using simple rules to be substantial.

The more experience people have with the task, the more they tend to solve it and generalize using information about individual stimuli (i.e. exemplars, Johansen & Palmeri, 2002). In particular, people use the perceived similarity between a representation of a novel stimulus and representations of previously observed stimuli to respond to the novel stimulus (Nosofsky, 1986). We consider the evidence that people generalize to unobserved stimuli using psychological similarity to stored representations to be strong.

4.3 Representational Change

While people carry out a task, representations of objects/tasks change over time. Here, we focus on a subset of three types of change mostly relevant for generalization. First, representations of individual stimuli become preciser with repeated exposure (Goldstone & Steyvers, 2001; Thalmann et al., 2023). Second, repeated exposure to the same sequences of items (e.g., F - B - I) leads to the emergence of chunks (Chase & Simon, 1973). These chunks are used in novel situations and tasks to deal more efficiently with limited working-memory capacity (Thalmann et al., 2019). Third, people learn more abstract representations, for example to categorize nouns as subjects and objects. They use sequential regularities on the abstract level (e.g., subject - verb - object) to generalize efficiently. That is, they handle previously unobserved sequences adhering to the same regularities with relative ease (e.g., Marcus et al., 1999; Wu et al., 2023). Whether regularities on an abstract level, however, are beneficial for generalization to novel tasks depends on whether the learned regularities can be found in other situations. This may not always be the case, especially outside of rule-based formal systems.

5 Cognitive and Motor Skill Training

How does the absence of generalization of training gains in the field of cognitive training relate to our proposed three stages? Transfer in the training literature is usually defined and measured as generalized process gains but not as representational generalization. For example, De Simoni and von Bastian (2018) argued that their participants had gained stimulus expertise without a change in the efficiency of the cognitive processes. We would assume that this representational change helped their participants to perform well in tasks requiring similar stimuli. For example, a learned three-digit chunk should be helpful in a different memory task with digits as stimuli. However, participants should perform again around baseline when they are transferred to a task requiring the same cognitive process to be carried out with completely new stimulus material, because they cannot profit from their acquired representations. In essence, we predict that generalization is a function of representations, which is diametrically opposed to the definition in the training studies.

The presence of generalized benefits in fitness and gross motor coordination contrasts with the cognitive training findings. A possible explanation is that the standardized tasks (i.e., standing broad jump and endurance shuttle run test) measure skills, which are already used in the individually trained disciplines. These tasks measure therefore training success, not generalization. The absence of generalization in the extreme case of the especial skill, can be seen as an extreme representational change. Increased representational precision for a particular scene (e.g., distance and angle of a basket, see panel C in Figure 2) is at the same time associated with decreased generalization given that scene. That is, already small changes in distance to the basket appear differently to the proficient basketball player. Perhaps surprisingly, the potential to use an object representation for generalization decreases with extensive practice.

6 Conclusion

Humans have been pictured as generalization artists, especially when compared to deep neural networks. Here, we presented a more detailed view and reviewed cognitive generalization mechanisms. We argued that successful generalization depends on whether the task representation and the strategy to solve a task align with the way a problem is set up. If an experimenter sets up a complicated, multidimensional, non-linear task structure, people may end up solving the task with an incorrect simple rule, because they have a strong preference for such simple rules; or they require thousands of training examples and feedback to eventually approximate the complicated function and generalize accordingly. In contrast, when the task requires the execution of a one-dimensional simple rule, people are likely to immediately perform and generalize well, because they can use their lifelong experiences to quickly constrain the problem space. Therefore, the true art of human generalization does not lie in a universal ability to generalize - but in the well-formed craft to establish and revise efficient representations in the face of limited processing capacity.

7 Recommended Reading

1. Taylor et al. (2021): Provides an extensive review of generalization phenomena with a particular focus on neurobiology.

2. Shepard (1987): Shows that the mapping from representations/descriptions of objects in physical space to psychological space is important when evaluating generalization in absolute identification tasks.

3. Nosofsky (1986): Explains how the same principles of exemplar-based processing are able to explain generalization in absolute identification and category learning.

4. Johansen and Palmeri (2002): Shows that people initially use rules to generalize to unobserved objects, but later on change to exemplar-based processing.

5. Cusack et al. (2024): Argues that human infants learn a generalizable foundation model in their protracted helplessness period, similar to self-supervised learning in natural language models.

8 Appendix

8.1 Funding

This work was funded by the Institute for Human-Centered AI at the Helmholtz Center for Computational Health.

Acknowledgements

References

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

Chase, W. G., & Simon, H. A. (1973). Perception in chess [Number: 1]. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

Cusack, R., Ranzato, M., & Charvet, C. J. (2024). Helpless infants are learning a foundation model [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *0*(0). https://doi.org/10.1016/j.tics.2024.05.001

De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, *147*, 829–858. https://doi.org/10.1037/xge0000453

DeLosh, E. L., McDaniel, M. A., & Busemeyer, J. R. (1997). Extrapolation: The Sine Qua Non for Abstraction in Function Learning. *23*(4), 968–986.

Donkin, C., Newell, B. R., Kalish, M., Dunn, J. C., & Nosofsky, R. M. (2015). Identifying strategy use in category learning tasks: A case for more diagnostic data and models. [Publisher: US: American Psychological Association]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 933. https://doi.org/10.1037/xlm0000083

Feldman, J. (1992). Constructing perceptual categories [ISSN: 1063-6919], 244, 245, 246, 247, 248, 249, 250–244, 245, 246, 247, 248, 249, 250. https://doi.org/10.1109/CVPR.1992.223268

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633. https://doi.org/10.1038/35036586

Fransen, J., Pion, J., Vandendriessche, J., Vandorpe, B., Vaeyens, R., Lenoir, M., & Philippaerts, R. M. (2012). Differences in physical fitness and gross motor coordination in boys aged 6–12 years specializing in one versus sampling more than one sport. *Journal of Sports Sciences*, *30*(4), 379–386. https://doi.org/10.1080/02640414.2011.642808

Geirhos, R., Temme, C. R. M., Rauber, J., Schütt, H. H., Bethge, M., & Wichmann, F. A. (2018). Generalisation in humans and deep neural networks. *Advances in Neural Information Processing Systems*, *31*. Retrieved December 20, 2023, from https://proceedings.neurips.cc/paper_files/paper/2018/hash/0937fb5864ed06ffb59ae5f9b5ed67a9-Abstract.html

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. [Publisher: US: American Psychological Association]. *Journal of Experimental Psychology: General*, *130*(1), 116. https://doi.org/10.1037/0096-3445.130.1.116

Goldwater, M. B., Don, H. J., Krusche, M. J. F., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General*, *147*(1), 1–35. https://doi.org/10.1037/xge0000387

Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E., & Binz, M. (2024, February). Ecologically rational meta-learned inference explains human category learning [arXiv:2402.01821 [cs]]. https://doi.org/10.48550/arXiv.2402.01821

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., & Schulz, E. (2023, July). *Zero-shot compositional reinforcement learning in humans* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/ymve5

Johansen, M., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*(4), 482–553. https://doi.org/10.1016/S0010-0285(02)00505-4

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. https://doi.org/10.1017/S0140525X16001837

Little, D. R., & Shiffrin, R. M. (2009). Simplicity Bias in the Estimation of Causal Functions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule Learning by Seven-Month-Old Infants [Publisher: American Association for the Advancement of Science]. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Mason, A., Szollosi, A., & Newell, B. (2022, September). Learning the lie of the land: How people construct mental representations of distributions. https://doi.org/10.31234/osf.io/rdhv8

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Nam, A. J., & McClelland, J. L. (2023, March). Systematic human learning and generalization from a brief tutorial with explanatory feedback [arXiv:2107.06994 [cs]]. Retrieved December 22, 2023, from http://arxiv.org/abs/2107.06994

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. [Publisher: US: American Psychological Association]. *Journal of Experimental Psychology: General*, *115*(1), 39. https://doi.org/10.1037/0096-3445.115.1.39

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working Memory and Intelligence–Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/10.1037/0033-2909.131.1.61

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

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

Simons, J. P., Wilson, J. M., Wilson, G. J., & Theall, S. (2009). Challenges to Cognitive Bases for an Especial Motor Skill at the Regulation Baseball Pitching Distance. *Research Quarterly for Exercise and Sport*, *80*(3), 469–479. https://doi.org/10.1080/02701367.2009.10599585

Taylor, J. E., Cortese, A., Barron, H. C., Pan, X., Sakagami, M., & Zeithamova, D. (2021). How do we generalize? *Neurons, behavior, data analysis and theory*, *1*, 001c.27687. https://doi.org/10.51628/001c.27687

Thalmann, M., Schäfer, T., Theves, S., Doeller, C. F., & Schulz, E. (2023, December). Task Imprinting: Another Mechanism of Representational Change? https://doi.org/10.31234/osf.io/s98vm

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37–55. https://doi.org/10.1037/xlm0000578

Vermaercke, B., Cop, E., Willems, S., D'Hooge, R., & Op de Beeck, H. P. (2014). More complex brains are not always better: Rats outperform humans in implicit category-based generalization by implementing a similarity-based strategy. *Psychonomic Bulletin & Review*, *21*(4), 1080–1086. https://doi.org/10.3758/s13423-013-0579-9

Wu, C. M., Schulz, E., Garvert, M. M., Meder, B., & Schuck, N. W. (2020). Similarities and differences in spatial and non-spatial cognitive maps (D. Marinazzo, Ed.).

PLOS Computational Biology, *16*(9), e1008149. https://doi.org/10.1371/journal.pcbi.1008149

Wu, S., Thalmann, M., & Schulz, E. (2023). Motif Learning Facilitates Sequence Memorization and Generalization [Publisher: OSF]. Retrieved December 20, 2023, from https://osf.io/2a49z