Prècis for

Guided by Generalization and Uncertainty: a Theory of Human Learning and Exploration

Charley M. Wu

How do people navigate the vastness of real-world environments where it is not feasible to explore all possibilities? The study of human learning has made rapid progress in the past decades, from discovering the neural substrate of reward prediction errors (Schultz, Dayan, & Montague, 1997), to using similar principles to build artificial intelligence capable of beating the best human players in skillful games such as Go (Silver et al., 2016). Yet this line of research has primarily focused on learning through repeated interactions with the same small set of stimuli (Lake, Ullman, Tenenbaum, & Gershman, 2017). How are humans able to rapidly adapt to novel situations and learn from sparse examples?

My thesis seeks to understand how people explore complex and uncertain environments, where sample efficient learning is a direct outcome of proficient exploration. Whereas learning efficiency is typically studied through the exploration-exploitation dilemma (should you explore a new option or exploit what is already known to be good?), this distinction breaks down in large problem spaces, where it is not enough to know whether to explore or exploit — one also needs to know *where* to explore.

Rather than exploring blindly or randomly, human exploration is guided by inductive biases that support remarkably efficient learning. The main contribution of my thesis is a theory of how humans use generalization from previous experiences and representations of uncertainty to search efficiently in novel situations. Generalization is modelled using Bayesian function learning (Rasmussen & Williams, 2006), where functions represent candidate hypotheses describing how people map limited experiences to novel possibilities. For instance, learning how the amounts of each ingredient relates to the taste of a soup ("maybe this needs more pepper"), is an example of a function that can be learned from sparse data and used to predict the value of hypothetical outcomes ("how would it taste if I added some roasted garlic?")

This approach to generalization uses the simple principle that similar actions produce similar outcomes, where a kernel representation of similarity can describe both continuous domains and discrete graph representations. By representing a distribution over hypotheses, this framework captures the underlying uncertainty of predictions. Uncertainty provides a powerful tool for sample efficient learning, allowing exploration to be strategically directed towards more uncertain (and therefore more informative) options.

My results demonstrates how a single theoretical framework is able to capture human learning and exploration across a wide range of spatial, conceptual, and graph-structured domains. Formalized as a computational model, my theory outperforms a wide range of 30 alternative models in predicting out-of-sample search behavior, simulating human-like learning curves, and predicting human judgments. Taken together, this thesis presents a new paradigm for studying human learning, providing novel insights into how human biases can improve Bayesian optimization algorithms and uncovering the source of developmental differences between how children and adults explore.

Chapter 1: A theory of generalization and exploration

Early research on stimulus-response learning and categorization defined generalization as a function of similarity (Shepard, 1987), where similar stimuli produce similar responses (Fig. 1a). This principle of similarity-based generalization can be reformalized as a kernel function (Jäkel, Schölkopf, & Wichmann, 2008), providing a mathematical framework for describing similarity in both continuous feature spaces (Fig. 1b) and discrete graph-structured domains (Fig. 1c).

Rather than generalization about concepts or category membership (Tenenbaum & Griffiths, 2001), I use kernel similarity to model how people learn implicit value functions over the space of possible options. Kernels provide inductive biases encoding expectations about the smoothness of functions (i.e., the extent that similar inputs produce similar outputs), which are used in a Gaussian Process framework (Rasmussen & Williams, 2006) to make flexible predictions about the value of novel stimuli by extrapolating or interpolating from existing data (Fig. 2). These predictions are Bayesian, providing estimates about the underlying uncertainty, which I show later to be a crucial component for guiding efficient exploration towards both promising and informative outcomes.

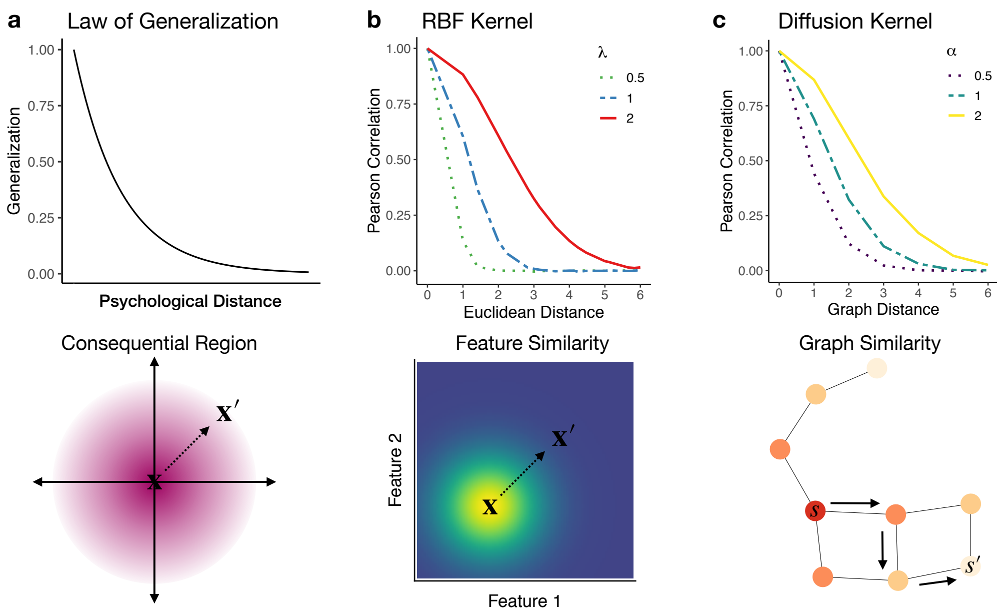

Figure 1. Theories of generalization. **a**) Shepard's (1987) law of generalization describes generalization as a function of distance between two stimuli $\mathbf{x}$ and $\mathbf{x}'$. Due to underlying uncertainty about the extent of a consequential region (pink), increasing distance between stimuli makes it less likely they belong to the same region. **b**) The radial basis function (RBF) kernel defines generalization on any feature space, where assumed correlation between any two points decays as a function of their distance, modulated by the length-scale λ . **c**) The diffusion kernel defines generalization on graphs, where larger graph distances (shortest path) typically correspond to lower assumed correlations of reward. The diffusion parameter α governs the rate of decay.

Chapter 2: Computational models

Chapter 2 provides a tutorial on the methods and computational models used throughout the thesis. I introduce reinforcement learning (RL; Sutton & Barto, 2018) as a computational framework for studying learning and exploration, where optimal solutions are only obtainable in limited cases and under restrictive assumptions. This motivates the need for learning strategies that explore efficiently.

I first describe common computational solutions for approximating the optimal solutions using tabular methods from Dynamic Programming, which still require exhaustive exploration of the search space. Since this scales poorly, an alternative approach is to use value function approximation to learn a global value function. These two computational approaches map onto two classes of models used to describe human learners. The option learning model (Fig. 2a) is based on classic theories of associative learning and resembles tabular methods by learning independent value representations for each option. The function learning model (Fig. 2b) is informed by theories of how humans explicitly learn functions (Lucas, Griffiths, Williams, & Kalish, 2015) and uses Gaussian Process regression as a method for generalizing previous observations of reward to a potentially infinite set of unexplored possibilities.

The option learning and function learning models are combined with a set of sampling strategies that transform the predictions of expected reward and the underlying uncertainty into a probabilistic choice framework (Fig. 3a). These sampling strategies represent different computational heuristics for navigating the exploration-exploitation dilemma, where I show that "optimism in the face of uncertainty" provides a remarkably accurate model of human exploration.

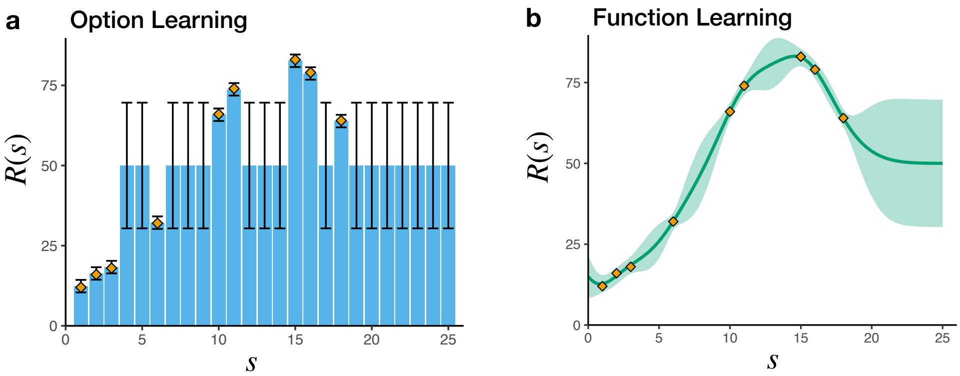

Figure 2. Models of Human Learning. **a**) The option learning model is a Bayesian mean tracker, which along with traditional models of associative learning, learn the expected rewards of each option (s) independently. Based on a set of observed rewards (orange diamonds) the option learning model makes predictions about the expected reward (blue bars) and the estimated uncertainty (error bars showing 95% CI). Unobserved options are defaulted to a prior mean and prior uncertainty. **b**) The function learning uses Gaussian process regression to generalize observed rewards (orange diamonds) across the search space, with the green line indicating expected rewards and the shaded ribbon indicating the uncertainty (95% CI).

Chapter 3: Exploration and generalization in vast spaces

The material in this chapter was published as Wu, Schulz, Speekenbrink, Nelson, and Meder (2017) and Wu, Schulz, Speekenbrink, Nelson, and Meder (2018).

To study how people search for rewards in large problem spaces, this chapter presents three experiments using multi-armed bandit problems with up to 121 options. Because the number of possibilities vastly outnumbered the available search horizon, performance heavily depended on the ability to generalize about unobserved options and guide exploration towards promising options. To provide traction for generalization, rewards were spatially correlated such that nearby options had similar rewards. Experiments 1 and 2 used generated environments with varying levels of spatial correlation, while Experiment 3 used natural agricultural datasets where rewards were defined by the yield of various crops (e.g., corn, wheat, and oranges), and spatial correlations arose organically, but heterogeneously both within and between environments.

Using a model Gaussian process model of generalization combined with an optimistic sampling strategy (Fig. 3a), I am able to predict participant behavior better than wide-range of alternative models and can simulate human-like learning curves. These results also reveal an intriguing bias towards undergeneralization. Through simulations, I show that undergeneralization can lead to better performance in a Bayesian optimization setting, demonstrating a crucial example of how cognitive science can inform computer science and machine learning.

Chapter 4: Learning like a child

The material in this chapter was published as Schulz, Wu, Ruggeri, and Meder (2019)

The richness of this spatial exploration paradigm also allows me to tackle an important question in developmental psychology, regarding the source of behavioral differences in how children and adults search for rewards. Children are prone to higher variability in their sampling behavior, with an influential hypothesis describing development as a "cooling off" process (Gopnik et al., 2017), where initially random behavior is reduced over the lifespan. Yet, a change in randomness is not the only explanation. High variability could also be attributed to higher levels of directed exploration, which is a more sophisticated sampling strategy that is informed by and directed towards uncertainty, rather than simply being more random. Additionally, differences in behavior could also be attributed to changes in cognitive processes and representations over the lifespan, for instance, in how children and adults learn and generalize, leading to different predictions about novel options.

These three theories of developmental differences correspond specifically to the different parameters of my model (Fig 3b). Random exploration is represented by the softmax temperature parameter, directed exploration is captured by the exploration bonus of upper confidence bound sampling, and the level of generalization is modeled using the length-scale of the Gaussian Process kernel. By developing a tablet experiment and testing subjects aged 7 to 55 in museums around Berlin, I investigated how generalization and exploration changes over the lifespan.

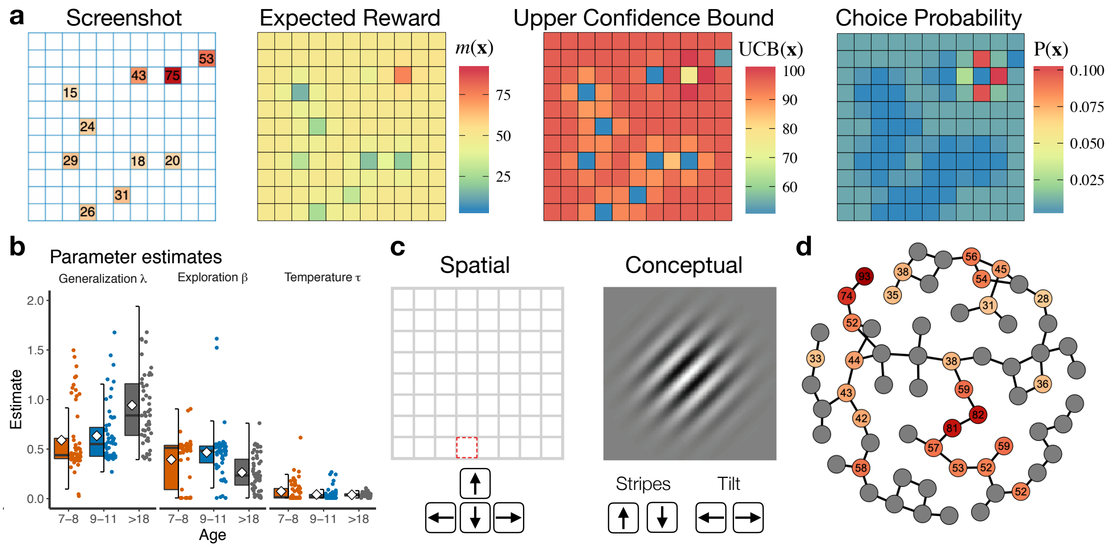

Figure 3. Overview of experiments. **a**) From left to right: the spatially correlated multi-armed bandit task, where participants search for rewards by clicking tiles on a grid and rewards are correlated by location. Expected rewards (uncertainty estimates not shown) as predicted by a Gaussian process function learning model, conditioned on the observations in the screenshot. The upper confidence bound of each option is a weighted sum of expected rewards and their estimated uncertainty, which can be interpreted as an "optimistic" inflation of value and provides a natural balance to the explore-exploit dilemma. A softmax choice rule transforms the upper confidence bound into probabilistic predictions about the next search decision. All plots use median parameter estimates from Experiment 2. **b**) Parameter estimates from Experiment 4, where the three parameters correspond to specific theories of developmental differences. The results show that children generalize less and explore more (directed towards uncertain options) than adults, but are not more prone to random, high temperature sampling. **c**) Stimuli from experiment 6 comparing learning in spatial and conceptual domains. Participants used the arrow keys to select a location in the spatial task or to modify a Gabor patch by changing the frequency of stripes and the tilt angle. Both tasks correspond to the same underlying reward distribution, where choices in similar locations or having similar Gabor features predict similar rewards. **d**) Graph correlated bandit task (Exp. 8), where transitions rather than singular feature values predict reward. Each node is a clickable reward generating option, with rewards correlated along the graph structure.

Converging evidence from behavior, computational models, and participant-based judgments indicate that children do not simply behave more randomly. Instead, children generalize less than adults, but explore the world more eagerly, in a distinctively directed fashion. This provides a richer picture of the developmental trajectory of learning, and suggest that to fulfill Alan Turing's dream of creating a child-like AI, we need to incorporate generalization and curiosity-driven exploration mechanisms.

Chapter 5: From spatial to conceptual search

The material in this chapter was published as Wu, Schulz, Garvert, Meder, and Schuck (2018)

Recent research has suggested that the same neural machinery is recruited to reason about both spatial and conceptual forms of knowledge, supporting the computation of distances or similarities between experiences (Constantinescu, O'Reilly, & Behrens, 2016; Garvert, Dolan, & Behrens, 2017). Yet, to what extent do spatial and conceptual reasoning share common computational principles, and in which ways are they different?

Chapter 5 presents two longitudinal experiments, where participants performed successive search tasks where either spatial or conceptual features predicted rewards. Experiment 5 used a paradigm where participants were shown both spatial and conceptual features (plant stimuli differing in the number of leaves and berries) simultaneous, but only one set of features predicted rewards. We found a task-order effect, where performance was boosted in the conceptual domain after experience in the spatial domain, but not vice versa. The parameters of the model were correlated across tasks, suggesting participants who generalized more or who explored more in one domain, also did so in the other.

Experiment 6 used an improved design where participants used arrow keys to modify a single stimuli displaying either spatial location on a grid or conceptual features (rotation and number of stripes of a Gabor patch; Fig. 3c). In both domains, the function learning model was the best predictor of search behavior, produced human-like learning curves, and was able to predict participant judgments. Yet there were also intriguing differences, where relative to the spatial domain, participants showed reduced levels of uncertainty-directed exploration and increased levels of random exploration in the conceptual domain. This supports the hypothesis that similar principles govern generalization across domains, but where the extended nature of our spatial awareness supports more sophisticated exploration strategies.

Chapter 6: Generalization in structured spaces.

The material in this chapters was published as Wu, Schulz, and Gershman (2019a) and Wu, Schulz, and Gershman (2019b).

This chapter extends the theories and models of the previous chapters from metric spaces to structured environments. From social networks to subway lines, many real-world human environments are defined by graph structures, where generalization is a function of the connectivity structure rather than the similarity of singular features. I use a diffusion kernel (Kondor & Lafferty, 2002) as a means to perform inference and learn functions over graph structures, where the RBF kernel used in previous chapters can be considered a special case. Thus, the diffusion kernel provides a broader framework of human generalization in both metric and structured spaces, with additional connections to neural theories of predictive coding (Stachenfeld, Botvinick, & Gershman, 2017).

Experiment 7 presents a function learning task where participants were asked to predict the number of passengers at various stations on virtual subway maps, based on observations at other stations. I show that the diffusion kernel accounted for both predictions and confidence judgments, beating out a variety of nearest-neighbor averaging heuristics. Experiment 8 was a bandit task similar to Experiments 1-6, but where the distribution of rewards were defined by the connectivity structure of a graph. The diffusion kernel best predicted participant choices, produced human-like learning curves, and could also predict judgments about expected reward and the underlying uncertainty of unobserved nodes.

Taken together, this paints a rich picture of how generalization is supported by a common organization of knowledge that is not only sensitive to spatial or feature similarity, but also encodes a predictive map of the transition dynamics that define structured environments.

Chapter 7: Conclusions

How do people make decisions beyond the limits of their experience? Rather than choosing blindly or randomly, I propose a theory of how human exploration is guided by generalization and uncertainty. People make predictive generalizations about novel options based on their similarity to previously encountered episodes. These predictions are then optimistically inflated based on the underlying uncertainty. Taken together, this provides a remarkably efficient learning strategy, where exploration is directed towards the most promising options.

Converging evidence from search behavior, judgments, and large-scale model comparisons provide robust support for my theory, which formalized as a computational model, is both recoverable and can simulate human-like learning. Across spatial, conceptual, and structured domains, this work finds common principles of learning and exploration with broad implications for understanding the nature of adaptive behavior in the face of uncertainty.

References

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*, 1464–1468.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, *6*, e17086.

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., ... Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, *114*, 7892–7899.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Similarity, kernels, and the triangle inequality. *Journal of Mathematical Psychology*, *52*, 297–303.

Kondor, R. I., & Lafferty, J. (2002). Diffusion kernels on graphs and other discrete input spaces. In *Proceedings of the 19th International Conference on Machine Learning* (pp. 315–322).

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*.

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*, 1193–1215.

Rasmussen, C. E., & Williams, C. (2006). *Gaussian Processes for Machine Learning.* MIT Press: Cambridge, MA.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science*, *30*(11), 1561-1572. doi: 10.1177/0956797619863663

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... others (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*, 1643–1653.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second ed.). MIT press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640.

Wu, C. M., Schulz, E., Garvert, M. M., Meder, B., & Schuck, N. W. (2018). Connecting conceptual and spatial search via a model of generalization. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1183–1188). Austin, TX: Cognitive Science Society.

Wu, C. M., Schulz, E., & Gershman, S. J. (2019a). Generalization as diffusion: human function learning on graphs. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (p. 3122—3128). Montreal, QB: Cognitive Science Society.

Wu, C. M., Schulz, E., & Gershman, S. J. (2019b). Searching for rewards in graph-structured spaces. In *Proceedings of the 2019 Conference on Cognitive Computational Neuroscience.* doi: 10.32470/CCN.2019.1041-0

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2017). Mapping the unknown: The spatially correlated multi-armed bandit. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1357–1362). Austin, TX: Cognitive Science Society.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*, 915–924. doi: 10.1038/s41562-018-0467-4