Sampling Heuristics for Active Function Learning

Rebekah A. Gelpi[1] (rebekah.gelpi@mail.utoronto.ca), **Nayan Saxena**[2] (nayan.saxena@mail.utoronto.ca),
George Lifchits[3] (glifchits@cs.toronto.edu), **Daphna Buchsbaum**[4] (daphna@brown.edu),
Christopher G. Lucas[5] (clucas@ed.ac.uk)

[1]Department of Psychology, University of Toronto
[2]Department of Statistical Sciences, University of Toronto
[3]Department of Computer Science, University of Toronto
[4]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University
[5]School of Informatics, University of Edinburgh

Abstract

People are capable of learning diverse functional relationships from data; nevertheless, they are most accurate when learning linear relationships, and deviate further from estimating the true relationship when presented with non-linear functions. We investigate whether, when given the opportunity to learn actively, people choose samples in an efficient fashion, and whether better sampling policies improve their ability to learn linear and non-linear functions. We find that, across multiple different function families, people make informative sampling choices consistent with a simple, low-effort policy that minimizes uncertainty at extreme values without requiring adaptation to evidence. While participants were most accurate at learning linear functions, those who more closely adhered to the simple sampling strategy also made better predictions across all non-linear functions. We discuss how the use of this heuristic might reflect rational allocation of limited cognitive resources.

Keywords: Function learning; active learning; sampling

Introduction

People must often learn and generalize from relationships between continuous quantities, where these relationships can take diverse forms. Temperatures rise and fall with the changing of seasons, trees grow steadily from saplings until they are fully mature, crops have a "sweet spot" of climatic constraints such as humidity and cold, foraging for food rests on contextual variables such as animal populations and water supply, and diseases can exponentially increase in the absence of constraints. In order to represent these relationships and use them to make accurate predictions, we must learn the underlying function from sparse observations to be able to predict unseen outcomes in a variety of new scenarios.

In addition to the general challenge of learning a function faithfully, most investigations of explicit function learning have focused on the human ability to use observed data to interpolate between previously observed points and to extrapolate beyond the limits of their experience. Given sufficient evidence, people can learn a wide variety of functional relationships (Bott & Heit, 2004; Lucas, Sterling, & Kemp, 2012; Wilson, Dann, Lucas, & Xing, 2015), but their inductive biases strongly favor linear relationships: people tend to learn linear relationships better than non-linear ones (Brehmer, 1974) and often extrapolate linearly even having learned that a relationship is non-linear in the data they have observed (DeLosh, McDaniel, & Busemeyer, 1997; Kalish, 2013).

Given this systematic bias towards linearity, learning non-linear relationships may require more or better evidence to overwhelm this strong *a priori* belief. Standard function learning experiments—at least those that have revealed the human ability to learn non-linear relationships—tend to rely on providing overwhelming evidence, such as multiple presentations of the same data point, large numbers of training examples, or multiple blocks of training (e.g., DeLosh et al., 1997; Kalish, Lewandowsky, & Kruschke, 2004; Kalish, 2013). This is at odds with the view that people are efficient learners, able to make good use of sparse evidence.

We hypothesize that difficulties with learning non-linear relationships may be, at least in part, an artifact of using passive observational designs—that is, designs in which participants do not choose which points to learn about. The evidence participants are presented with in these experiments is often of limited utility, with the most informative observations presented alongside a large number of comparatively unhelpful ones, increasing participants' attentional and memory burden, as well as fatigue. While active learning provides benefits beyond efficient selection of samples (e.g., Markant & Gureckis, 2014a; Markant, Ruggeri, Gureckis, & Xu, 2016), we focus here on testing whether people tend to choose useful samples, what policies may underlie the samples that people choose, and whether their sampling strategies facilitate better learning of linear as well as non-linear functions.

Previous work in active learning has suggested that people can effectively learn linear functions by focusing their sampling on regions of high uncertainty (Jones, Schulz, Meder, & Ruggeri, 2018). An uncertainty-based sampling policy could also be useful for non-linear functions, as maximizing information gain about the extrema of a function eliminates the need for extrapolation, which can otherwise be inaccurate in non-linear functions (DeLosh et al., 1997; Kalish et al., 2004). However, the computational demand of iteratively adjusting one's sampling strategy could mean that, especially in more complex non-linear domains, people trade off optimal behaviour against the cognitive or temporal costs of doing so (Gershman, Horvitz, & Tenenbaum, 2015). In these circumstances, using less accurate or flexible heuristics could nonetheless be *resource-rational*, or optimal under constraint (Gigerenzer, 2008; Lieder & Griffiths, 2020).

In the following sections, we first present a Gaussian process-based framework for representing the task of func-

tion learning, describing how a learner can use their existing knowledge about a functional relationship to interpolate or extrapolate to unknown function values. We then introduce simple candidate strategies that an active learner might use to select samples and an experimental task to assess humans' sampling behaviour and predictions, and compare these to simulated learning under different sampling policies.

Gaussian Process Model

We use a Gaussian process (GP) model (e.g., Griffiths, Lucas, Williams, & Kalish, 2008; Lucas, Griffiths, Williams, & Kalish, 2015) to provide a general framework for understanding both rule-based and similarity-based function learning. This model uses samples $\mathbf{x}_n = (x_1 \ldots x_n)$ to further approximate a learned function f by inducing a Gaussian distribution on the observed $y_i = f(x_i)$ values based on sampled x_i values (Rasmussen & Williams, 2006). For known function outputs f and new unknown points f_* the joint probability distribution is then defined as

$$\begin{pmatrix} f \\ f_* \end{pmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right)$$

In the above equation, we have $K = k(x,x)$, $K_* = k(x,x_*)$ and $K_{**} = k(x_*,x_*)$, where k denotes the generalized class of Matérn kernel given by:

$$k(x_i,x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{\ell} d(x_i,x_j) \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu}}{\ell} d(x_i,x_j) \right)$$

where d denotes Euclidean distance, K_{ν} is a modified Bessel function, and Γ is the standard Gamma function. Throughout this paper we use this model with smoothness $\nu = 1.5$ and length-scale $\ell = 0.1$ as these parameters provided consistent, natural and realistic function interpolations for varied distributions of known data and functional forms.

Sampling Strategies

Under this GP framework, as we get further from previously-observed points, our uncertainty increases, which could lead to more heterogeneous or inaccurate extrapolation if a function's minimum and maximum values are not known. Therefore, for a learner with limited opportunities to sample, we expect that the most effective and informative strategies will include sampling the extrema. We assess the informativeness of three policies: uniform random sampling, an equidistant sampling policy that selects the minimum and maximum values and interpolates equally between them, and uncertainty-based sampling.

Random sampling   One simple strategy to learn about a function is to sample randomly from the possible domain of values. While random sampling is straightforward, decision-makers may display substantial sub-optimality in their sampling choices if they happen to sample multiple points in

close proximity that are unlikely to be maximally informative. Given human inductive biases towards linearity, this could result in difficulty extrapolating or interpolating non-linear functions where no samples have been drawn. We represent this policy as drawing samples from a distribution where each sample $x_i \sim \text{Uniform}(0,1)$.

Equidistant sampling   As mentioned previously, sampling the minimum and maximum feasible values within the confines of the problem to be solved minimizes the need for extrapolation, and partitioning the remaining space among the remaining samples likewise balances interpolation between the remaining points, making it suitable for relatively accurate prediction across many commonly-encountered functions. While somewhat inflexible, this may represent a highly tractable, efficient heuristic for sampling within a limited domain. We represent this sampling approach as taking the N samples to be drawn and allotting them in such a way that the points sampled roughly reflect $N-1$ equal partitions of the space to be sampled:

$$x_i \sim \text{Beta}\left(1 + \frac{i\rho}{N-1}, 1 + \frac{(N-1-i)\rho}{N-1} \right) \; (i \in \{0 \ldots N-1\})$$

To reflect a moderate preference for sampling from the peaks of the beta distributions while allowing for some samples to be drawn from nearby values, we chose a free parameter value of $\rho = 10$ that denotes the size of the peaks of the sampling distribution.

Adaptive sampling   The two above algorithms reflect sampling procedures that do not dynamically update based on new information. In order to choose the best samples, however, it may be effective to adapt one's strategy to account for already sampled points. Given that people preferentially draw samples in regions that resolve current uncertainty, even when this does not maximize information gain (Markant & Gureckis, 2014b), we designed a sampling algorithm motivated by a myopically ideal strategy that adapts to new information by identifying the point of highest uncertainty.

This sampling procedure uses a GP model that greedily chooses points from the domain-space by iteratively fitting the model on the previously sampled points, and choosing the next sample as the point on the posterior distribution of GP functions with the highest variance.

To test the effectiveness of these algorithms and compare their performance to human sampling strategies as well as the learned functions, we designed an experimental task in which participants must select a small number of samples to try to optimize their performance in a prediction task.

Experimental Design

Recruitment and Procedure   98 adult participants ($M_{\text{age}} = 31.8, SD_{\text{age}} = 11.1$) were recruited through Prolific and paid £1.00 for completing an online learning task presented via a web-based program.

Familiarization and Exposure   In the task, participants were told that they would be playing the role of a scientist. The scientist's job was to learn about a number of possible drugs, each being researched for their role in improving patients' well-being. Ultimately, the goal of the participant was to learn the relationship between the length of time that a drug was provided to patients and the patient's well-being scores.

Before participating in the experimental trials, participants were familiarized with a warm-up trial where they were presented with two empty horizontal bars. The first bar was labelled *orange juice consumed* and the second bar was labelled *hours of sleep*. Participants were told that they could drag the first bar to change the amount of orange juice consumed. After confirming their choice, they learned about the amount of sleep of an individual who consumed the specified amount of orange juice, which is displayed on the second bar. Next, participants were shown a specified amount of orange juice consumed, and asked to predict the amount of sleep on the second bar by dragging the bar to the predicted amount.

Experimental Task   After completing the familiarization trial, participants completed the experimental task. Participants learned about four total drugs, presented in a randomized order of four blocks of trials. Each drug had a predefined relationship to the well-being of the patient: positive linear ($y = 0.8x + 0.1$), exponential ($y = 100^{(x-1)}$), quadratic ($y = 0.95 - \frac{(x-0.5)^2}{0.3}$), or periodic ($y = 0.5\sin(7x) + 0.5$). For each drug, participants completed a block of trials with the same format as the warm-up. They first sampled time points since the patient began taking the drug, to learn about its effect on well-being at a given time point, by manipulating a blue horizontal bar (Figure 1). After sampling 5 points to learn about, participants were asked to predict the well-being of a patient at 8 randomly presented pre-selected time points (values of $x \in \{0.01, 0.15, 0.29, 0.43, 0.57, 0.71, 0.85, 0.99\}$), displayed on the blue bar. Participants were not given the exact values of data points, so the domains of x and y presented here are arbitrary.

Figure 1: Participants chose 5 points to learn about the drug's effect on well-being by changing the value of the blue bar (top), and then made 8 predictions using the red bar (bottom).

Results

Simulations   To analyze the effectiveness of the identified sampling strategies and establish comparison baselines, we generated synthetic data for 100 participants under each sampling strategy: random, equidistant, and adaptive sampling, and fitted Gaussian process models for each function (Figure 2) using the `sklearn` library in Python. The correlation coefficient and the pooled and average mean squared error (MSE) of the GP fits were then computed and compared (Table 1).

Sampling randomly (row 1) resulted in reasonably good approximations of the true function at the centre but demonstrated wider deviations at the minimum and maximum values. As a large number of the simulated participants failed to sample from the extrema of the function, this led to more inaccurate extrapolations in areas where no samples were drawn. The pooled and average MSE of simulations of this policy demonstrates a comparative disadvantage at approximating the functions when compared to equidistant and uncertainty minimization sampling.

In contrast, the process of sampling approximately equidistant x values (row 2) appeared to perform better at producing an estimate for the true functions. A considerable improvement in terms of the fit of the Gaussian process model can be visibly noticed as compared to random sampling. Notably, equidistant sampling generated the lowest pooled and average individual MSE of the three sampling strategies.

Finally, similar to the equidistant heuristic strategy, the adaptive sampling algorithm (row 3) frequently sampled from regions in the neighbourhood of zero and one. Overall, the adaptive sampling strategy yielded comparable, though slightly higher, MSE values to the equidistant strategy.

Given the cognitive costs of adaptive sampling, even when calculated approximately rather than through the representation of the full posterior distribution of candidate functions, we do not expect people to utilize this strategy to draw samples. Nevertheless, the similarity of this adaptive strategy to a simple heuristic strategy of sampling evenly spaced points in addition to the minimum and maximum values, suggests that if people use such a strategy, they will select comparably informative points to learn about, improving their subsequent predictions of the true function's value.

Sampling Task   Gaussian process models were also fitted on samples for each of the 98 human participants across all four functions. These models were then compared against previously established baselines produced by fitting GP models on generated data from different sampling strategies.

As observed in Figure 2 (row 4), the models fitted on human samples were able to predict the true function and were comparable to the equidistant or adaptive sampling strategies in terms of pooled approximation error (MSE), although people's choices showed considerable variability and some individuals chose less informative points to learn about; for example, 15 participants did not sample a value above 0.5 for at least one of the functions. As a result, the average
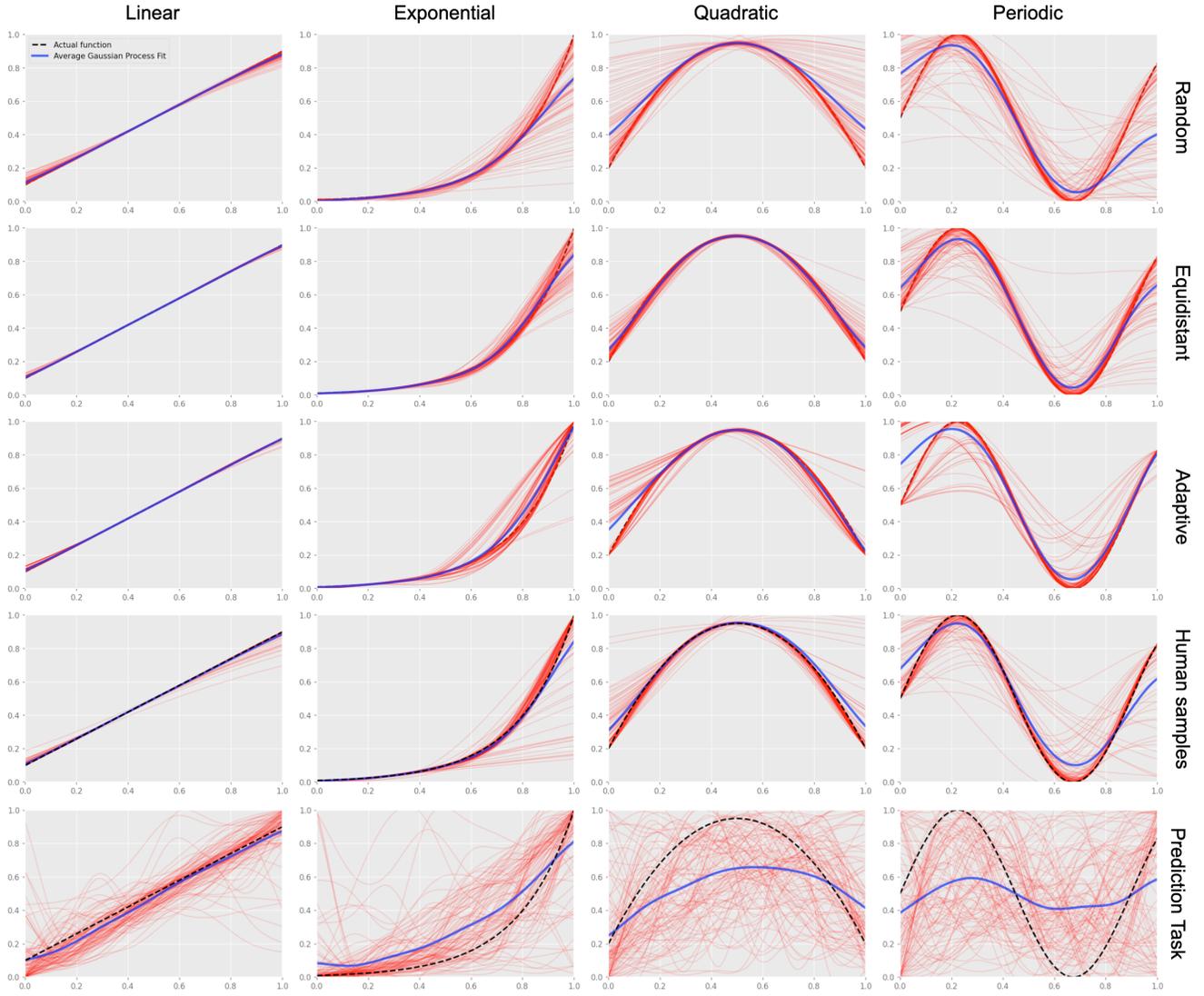
Figure 2: Gaussian process best fits for random sampling (row 1), equidistant sampling (row 2), adaptive sampling (row 3), human samples (row 4), and human predictions (row 5). Red lines represent individual GP fits for samples with a given sampling strategy or fits for individual learned functions on the prediction task. The blue line represents the mean of all GP fits.

Function	Sampling strategy	Pooled MSE	Average MSE	r
Linear	Random	$3.97 \cdot 10^{-5}$	0.0001	> 0.999
	Equidistant	$1.63 \cdot 10^{-7}$	$6.88 \cdot 10^{-6}$	> 0.999
	Adaptive	$4.87 \cdot 10^{-6}$	$2.23 \cdot 10^{-5}$	> 0.999
	Human	$2.18 \cdot 10^{-5}$	0.011	> 0.999
Exponential	Random	0.003	0.008	0.991
	Equidistant	0.001	0.002	0.995
	Adaptive	0.001	0.004	0.999
	Human	0.001	0.044	0.998
Quadratic	Random	0.044	0.015	0.995
	Equidistant	0.0004	0.002	0.997
	Adaptive	0.002	0.005	0.989
	Human	0.002	0.099	0.996
Periodic	Random	0.014	0.046	0.940
	Equidistant	0.002	0.009	0.995
	Adaptive	0.005	0.023	0.993
	Human	0.006	0.154	0.980

Table 1: Pooled MSE, average individual MSE, and correlation coefficients for sampling strategies and human samples.
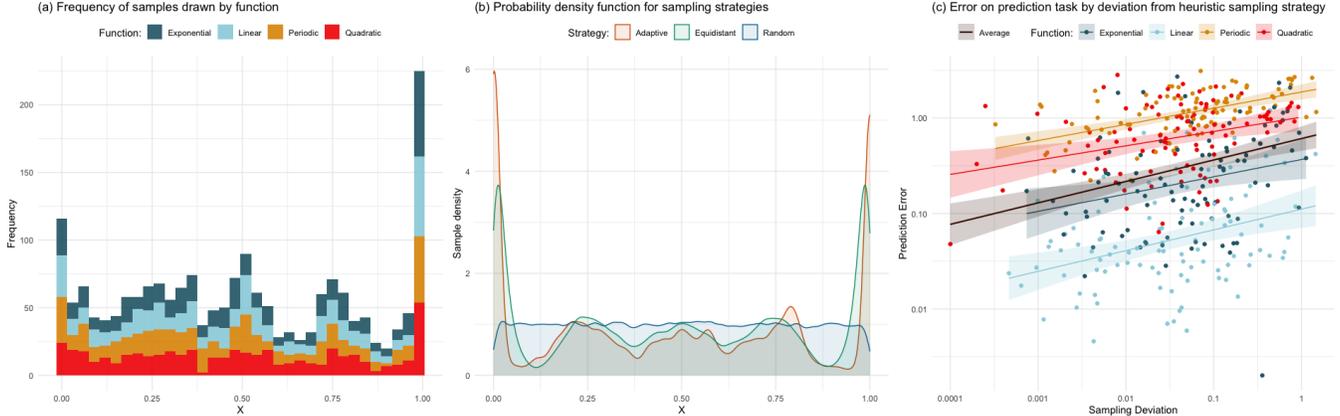
Figure 3: (a) Histogram showing frequency of human sampled points across the x axis by function; participants' sampling behaviour was comparable across all four function families. (b) Estimated probability density functions for the random (blue), equidistant heuristic (green), and adaptive (orange) sampling strategies. (c) Participants were most accurate in predicting values for the linear function (light blue), followed by the exponential (dark blue), quadratic (red), and periodic (orange), in order, and were more accurate when their choices closely matched the equidistant strategy.

Human sampling (vs.)	Equidistant	Random	Adaptive
KL divergence (D_{KL})	**0.052**	0.077	0.103
Hellinger distance (H)	**0.113**	0.140	0.161

Table 2: Distance metrics for sampling strategies.

MSE for individual fits was relatively high. Nevertheless, in aggregate, the probability density of participants' sampled choices showed the greatest similarity to the equidistant heuristic model predictions, compared to the random sampling and adaptive sampling strategies (Table 2), as measured by Kullback-Leibler divergence (a measure of difference between probability distributions) and Hellinger distance (a measure of similarity between distributions).

Prediction Task To measure participants' prediction errors as well as their sampling behaviour, we calculated the sum of squared errors of participants' predictions from the true function, as well as the deviation of their samples from the predictions of the equidistant heuristic strategy. As participants' errors for both the prediction task and the samples were not normally distributed, sum of squared error (SSE) scores were scaled and log transformed before performing inference.

We first tested the hypothesis that prediction accuracy would be correlated with adherence to the equidistant sampling heuristic by running a Bayesian mixed-effects linear regression, with function and sampling deviation as predictors for prediction accuracy, and random intercepts for baseline accuracy, using the `brms` package in R, with uninformative priors of $\mathcal{N}(0,3)$ placed on coefficients and a zero-truncated Cauchy$(0,2)$ prior placed on the standard deviation.

Confirming prior findings on the difficulty of learning non-linear relationships, participants were best at the linear, with increasing error, in order, on predictions for the exponential function ($\beta = 0.87$, 95% CI: [0.71, 1.04]), followed by the quadratic function ($\beta = 1.61$, 95% CI: [1.45, 1.77]) and the periodic function ($\beta = 1.96$, 95% CI: [1.81, 2.12]) relative to baseline performance on the linear function (Figure 3c).

Across all functions, there was strong evidence that greater deviation from sampling points consistent with the equidistant heuristic was associated with greater prediction errors, ($\beta = 0.19$, 95% CI: [0.12, 0.26]).

Further, while participants in the quadratic and periodic conditions had greater errors on average, this did not appear to be because participants were failing to infer high-level features of the functions such as their non-monotonicity. To investigate this phenomenon, we calculated the standard deviation of interpolated slopes between participants' predicted points as a measure of the variability of the predicted function. Higher slope variability reflects larger deviation from a constant slope, such as a rising and falling function, while lower slope variability reflects a constant slope, closer to a flat or linear relationship. There was strong evidence that participants had more slope variability in predicting periodic ($\beta = 1.38$, 95% CI: [1.00, 1.75]) and quadratic ($\beta = 0.93$, 95% CI: [0.71, 1.15]) functions. Participants with lower variability made fewer errors in predicting the linear function ($\beta = 0.93$, 95% CI: [0.68, 1.16]), but no such relationship existed for the periodic ($\beta = -0.38$, 95% CI: [−0.73, 0.04]) or quadratic ($\beta = -0.10$, 95% CI: [−0.39, 0.19]) functions.

Participants making predictions for non-linear functions may have learned broader features of the functions, such as the fact that the function had both increases and decreases at various points, while not necessarily encoding exact values or appropriate parametrizations for the functions, confirming prior findings that participants readily extrapolate non-linear functions even in sparse environments (León-Villagrá, Preda, & Lucas, 2018). Failure to encode exact values while retaining some qualitative representation of the higher-order structure of the function might also explain why participants disproportionately estimated the minimum x value of the function to likewise have a y value of zero, as participants may have relied more heavily on inductive biases for an intercept value of zero (Kwantes & Neal, 2006).

General Discussion

Despite strong inductive biases towards positive linear relationships, people are able—at least in aggregate—to learn a variety of functional relationships, even when given very sparse evidence, and appear to be able to apply a relatively simple heuristic strategy, sampling the minimum and maximum x values and evenly spaced points in between, that requires little cognitive effort while providing a comparable outcome, and comparable information gain, to a more computationally expensive adaptive strategy. We also found that those who made choices more in line with this policy were more accurate in the prediction task.

Respondents showed vast variability in the informativeness of their samples and the accuracy of their predictions; however, aggregated results showed that a recognizable parametrization of the true function was learned, perhaps the product of a "wisdom of crowds" effect (Steyvers, Miller, Hemmer, & Lee, 2009) averaging out individual errors. Figure 2 shows the averages are compressed toward the center of the range relative to the true function, which is expected if some judgements are corrupted by additive zero-mean noise that is truncated at the limits of the values participants can select, or sometimes selected uniformly at random. Nevertheless, as Wilson et al. (2015) have pointed out, averaged responses can eliminate important statistical structure in human predictions of functional relationships, and our analysis revealed that even inaccurate individuals' predictions suggested that they had learned, for example, when a function was non-monotonic.

While the proposed sampling strategy we introduced is relatively inflexible, this policy could reflect the use of rational metareasoning (Lieder & Griffiths, 2017), with participants deploying a heuristic with a favourable trade-off between its utility in giving relatively informative evidence for a variety of common functional relationships (including the most *a priori* plausible, positive linear), while requiring little cognitive effort to adapt to existing sampled points. This also coheres with previous findings in active function learning, where participants' choices most closely fit a simpler linear regression policy rather than a generalized GP policy when learning in linear domains (Jones et al., 2018).

In this view, the equidistant heuristic may not be deterministically employed in all situations in which people must choose limited information to learn about a relationship, but could perhaps be used situationally in a rational way. Future research could place learners in a situation where use of a similar heuristic would lead to less informative evidence; if this heuristic is a rationally-adapted strategy trading off accuracy and efficiency, then we predict learners would adapt and deploy a different strategy for learning on such a task. Nevertheless, for limited domains, simple heuristics such as the one we have outlined may be a valuable element of the human cognitive toolkit to approximate optimal learning strategies.

References

Bott, L., & Heit, E. (2004). Nonmonotonic Extrapolation in Function Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 38–50.

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.

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), 968–986.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.

Griffiths, T. L., Lucas, C. G., Williams, J. J., & Kalish, M. L. (2008). Modeling human function learning with Gaussian processes. In *Advances in Neural Information Processing Systems* (Vol. 21, p. 553–560).

Jones, A., Schulz, E., Meder, B., & Ruggeri, A. (2018). Active function learning. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 578–583).

Kalish, M. L. (2013). Learning and extrapolating a periodic function. *Memory & Cognition*, *41*, 886–896.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of Linear Experts: Knowledge Partitioning and Function Learning. *Psychological Review*, *111*(4), 1072–1099.

Kwantes, P. J., & Neal, A. (2006). Why people underestimate y when extrapolating in linear functions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1019–1030.

León-Villagrá, P., Preda, I., & Lucas, C. G. (2018). Data Availability and Function Extrapolation. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 2017–2022).

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762–794.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*.

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*(5), 1193–1215.

Lucas, C. G., Sterling, D., & Kemp, C. (2012). Superspace extrapolation reveals inductive biases in function learning. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 713–718).

Markant, D. B., & Gureckis, T. M. (2014a). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94–122.

Markant, D. B., & Gureckis, T. M. (2014b). A preference for the unpredictable over the informative during self-directed learning. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 958–963).

Markant, D. B., Ruggeri, A., Gureckis, T. M., & Xu, F. (2016). Enhanced Memory as a Common Effect of Active Learning. *Mind, Brain, and Education*, *10*(3), 142–152.

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

Steyvers, M., Miller, B., Hemmer, P., & Lee, M. D. (2009). The Wisdom of Crowds in the Recollection of Order Information. In *Advances in Neural Information Processing Systems* (pp. 1785–1793).

Wilson, A. G., Dann, C., Lucas, C. G., & Xing, E. P. (2015). The Human Kernel. In *Advances in Neural Information Processing Systems* (Vol. 28).