Can Children Learn Functional Relations Through Active Information Sampling?

Caiqin Zhou[1](caiqin_zhou@brown.edu), **Rebekah A. Gelpí**[2,3] (rebekah.gelpi@mail.utoronto.ca),
Christopher G. Lucas[4](clucas2@inf.ed.ac.uk), & **Daphna Buchsbaum**[1] (daphna_buchsbaum@brown.edu)

[1]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University
[2]Department of Psychology, University of Toronto [3]Vector Institute
[4]School of Informatics, University of Edinburgh

Abstract

Functional relations are prevalent in everyday life and science. Do children have intuitive knowledge of functional relations, and can they learn these relations by active information gathering (i.e., choosing a few input values and observing the corresponding outputs)? We found that 6- to 9-year-olds can learn different families of functions (linear, Gaussian, and exponential) through both informative data provided by an experimenter and data they gather from the environment for themselves. Overall, children learn linear functions more accurately than non-linear functions. When choosing data points to learn about, some children select highly similar points that only shed light on a narrow region of a function, while others choose more variable inputs and gain a more holistic view of a function. Children who use this latter, globally informative strategy have higher learning accuracy, particularly for non-linear functions. Results suggest that children are in the process of developing effective strategies for active function learning.

Keywords: function learning; active learning; sampling

Introduction

Imagine that a child wants to make playdough with the perfect texture. Although online recipes can provide a general ratio of different ingredients (e.g., water, flour, oil, and cream of tartar), the child may still need to find out the specific combination that works best for them. For example, underhydrated dough will be crumbly, over-hydrated dough will be sticky, and the optimal water-to-flour ratio may depend on the home humidity level. Trying numerous ratios and kneading each dough is time-consuming, so the child may judiciously choose a few ratios to test out to determine the best recipe.

Function learning describes this process of learning the relations between inputs and outputs and using them to make predictions about novel scenarios. The ability to learn functional relations is crucial to everyday life and science, whether the relation to be learned is how hydration levels affect playdough texture or how infectious diseases spread over time. Besides the difficulty of inferring the functional form and parameters accurately in the vast space of possibilities, function learning in real life is challenging because learners often do not passively receive the relevant information. Just as determining the relation between hydration levels and playdough texture requires choosing various ingredient ratios to test out, function learning often requires active selection of useful data, a process referred to as active learning.

In the present study, we explore how children learn functional relations through self-directed information search. Is knowledge about functions part of children's intuitive understanding of the world, or does it emerge later, perhaps through formal math education? Much of what we know about children's function learning comes from education research that has focused on ways to teach functions through math lessons.

Education studies suggest that, with appropriate lessons, elementary school children can learn simple functions (Blanton & Kaput, 2004; Blanton et al., 2015; Stephens et al., 2017). Using data from everyday scenarios (e.g., a number of dogs and the corresponding number of total eyes), students can learn linear relations and use them to predict outputs for novel input values as early as first grade, and they reliably do so from third grade on. However, these lessons require reasoning with exact numbers and verbalizing one's thinking. Causal learning and scientific thinking studies show that children succeed on implicit reasoning tasks before they succeed on explicit versions of the same tasks (Shtulman & Walker, 2020; Weisberg & Sobel, 2022). Thus, children may show an earlier understanding of functions in intuitive reasoning tasks.

Function learning in non-classroom contexts has been widely studied in adults, typically by showing learners numerical input-output pairs before asking them to predict outputs for new input values. Adults can learn various functions (e.g., linear, quadratic, cyclic; Bott & Heit, 2004; Wilson et al., 2015), but their inductive biases favor learning linear functions (Brehmer, 1974; McDaniel & Busemeyer, 2005) and they often extrapolate linearly, even when the true relation is non-linear (DeLosh et al., 1997; Kalish, 2013). Recent studies suggest that adults can also learn functions when input-output pairs are presented graphically in scatterplots (E. Schulz et al., 2017), and they learn linear functions more accurately than non-linear (exponential) ones in these visual trend detection tasks as well (Ciccione et al., 2022).

Only a few studies have tested children's intuitive knowledge of continuous mathematical functions. Coates et al. (2023) presented functions to 4- to 7-year-olds in causal scenarios, in which flashing lights made flowers bloom. Children can match the lights and flowers by the functions that underlie their change (e.g., rates of flashing or blooming). They can distinguish functions both across and within function families (monotonic, U-shaped, and periodic) and can match visualizations to descriptions. Notably, this study has tested comprehension using matching decisions based on contrasting options, but function learning requires more than being able to distinguish functions, such as the ability to interpolate between and extrapolate beyond observed data points.

Another set of studies tested children's fine-grained un-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

derstanding of linear and exponential functions by showing them the initial steps of a growth curve before asking them to forecast the subsequent steps. Across these studies, 5 to 13-year-old children predicted linear growth accurately. When children begin to understand exponential functions, however, is less clear (Ebersbach & Wilkening, 2007). Five-year-olds distinguished linear and exponential functions to some extent by predicting distinct growth rates, yet both rates were constant (Ebersbach et al., 2010). Nine-year-olds' predictions of exponential growth followed the correct curve in one study (Ebersbach & Resing, 2008) and a straight line in another study (Ebersbach et al., 2008). This evidence suggests that children, like adults, may show a linear bias, and that children's understanding of non-linear functions may emerge between kindergarten and the early elementary school years.

These past studies have relied on passive observational designs, in which children learn functions from data provided by an experimenter. However, function learning in the real world – as in our playdough example – often involves first selecting what data to sample and then drawing appropriate inferences based on those data. Data gathering is often costly, as opportunities to sample data are limited and the process of data gathering is time-consuming. To ensure that the limited opportunities for information gathering are well-used, learners need to judiciously choose the most useful data to sample.

Active learning studies have examined how children and adults learn reward distributions when searching in environments with spatially-correlated rewards, as learners balance needs for reward and information (e.g., Meder et al., 2021; Wu et al., 2018). However, we focus on how people learn functions when their search is solely driven by the need for information. Recent studies show that adults can learn various functions (e.g., linear, exponential, quadratic, cyclic) by active data gathering. They use equidistant sampling (i.e., sampling the minimum and maximum of all possible inputs and evenly spaced values in between), instead of computationally costly strategies that involve adaptation to newly sampled data (Gelpi et al., 2021, 2023). This strategy reduces the need for extrapolation (i.e., predicting outputs for inputs outside the range of observed input values), which is a particularly difficult aspect of function learning, especially for non-linear functions (DeLosh et al., 1997; Kalish et al., 2004).

Active learning studies have also shown powerful ways in which children can learn through their own actions, for example, in causal contexts (Sim & Xu, 2017; Sobel & Sommerville, 2010). Preschoolers selectively explore when they encounter inconsistent or confounded information, and exploration allows them to resolve the ambiguity (Legare, 2012; L. E. Schulz & Bonawitz, 2007). Yet if children can actively learn mathematical functions is less explored. In the current study, we examine if children can learn different forms of functions through data they gather for themselves, in addition to informative data provided to them. We also characterize the information-seeking strategies children may use.

Our task involved learning about different growth patterns, as children understood functions in the context of growth (e.g., Coates et al., 2023) and growth curves could follow various shapes. Children learned about apple growth over time in an orchard. They first saw data on apple quantities at a few time points before making predictions for unobserved time points. Depending on the trial type, the initial data were either chosen by the experimenter (Prediction Only trial) and highly informative, or chosen by the children themselves (Sample + Prediction trial). Specifically, the informative data were equidistant samples that generally led to accurate function learning in adults (Gelpi et al., 2021, 2023). We expected some children might use this effective, adult-like strategy in their own data gathering, sampling the extrema of a function and perhaps evenly-spaced points in between. Children who are less systematic in their search might densely sample one area of the function domain. We predicted that children whose samples are dispersed might gain a more holistic and accurate view of the function than those with clustered samples that only shed light on a small section of the function.

Although learning from self-generated data enhances children's memory (Markant et al., 2016) and allows them to make more accurate causal inferences (Sobel & Sommerville, 2010) than passively viewing the same data, the benefits of active learning may not all transfer to the current task. In our study, self-selected data are pitted against highly informative data given by the experimenter (instead of data chosen randomly or by a yoked peer); if children do not yet gather data effectively, these benefits may be counteracted by a lack of useful data to learn from. We therefore have no directional hypothesis on how children may learn differently from self-selected data versus passively received useful data. Our goal is to examine if children can learn various functions both with maximal scaffolding and through self-directed exploration.

We focus on three function families: linear, Gaussian, and exponential. These functions vary in abstract features (linearity and monotonicity), allowing us to test if children's learning is sensitive to these features. Since children find these functions increasingly hard in math class (Stephens et al., 2017) and understand linear relations earlier than exponential ones (Ebersbach & Wilkening, 2007), we expect that children may learn linear functions better than non-linear ones. Alternatively, as children have more diffuse priors than adults (Lucas et al., 2014), they may not find non-linear functions harder than linear ones. We focus on 6- to 9-year-olds, as children's knowledge of non-linear functions seems to emerge between kindergarten and the early elementary school years (Ebersbach & Resing, 2008; Ebersbach et al., 2008, 2010).

Methods

We investigated if children can learn functional relations and use them to predict outputs for new input values, (1) through passively observing informative data selected for them by an experimenter (Prediction Only trials) and (2) through self-selected data (Sample + Prediction trials). We also characterized children's strategies for self-directed data gathering.

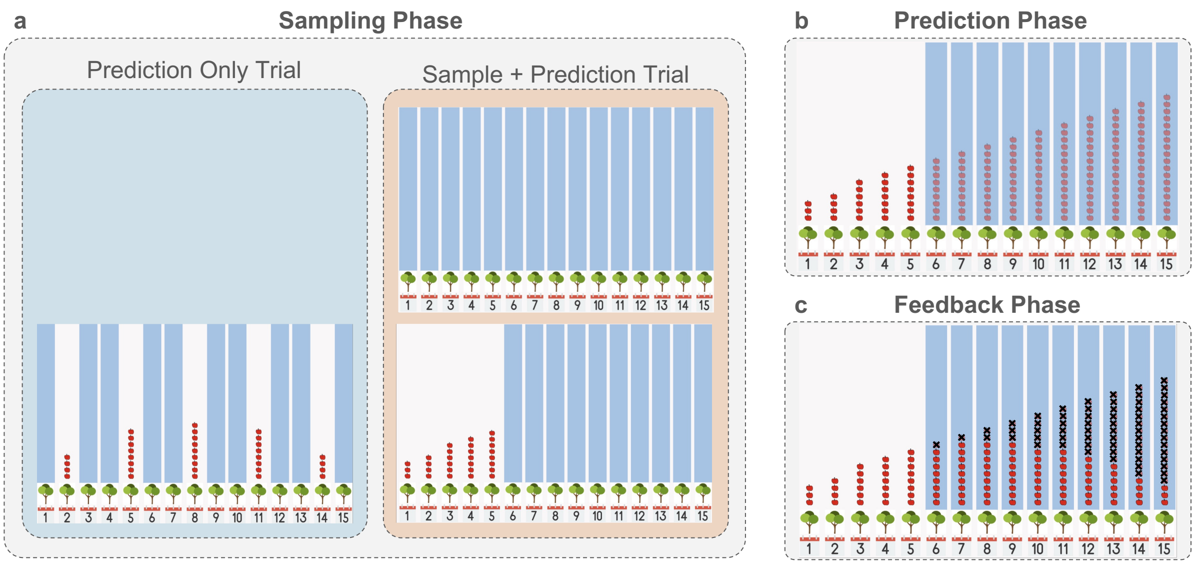

Figure 1: Example trial (true function: Gaussian). (a) Sampling phase. The timeline represented a 15-day period, and the apples above each day showed the number of apples that grew that day. In Prediction Only trials, the apple quantities for five days were already shown. In Sample + Prediction trials, children sampled five time points, by tapping the blue bar above a given day to reveal the apple quantity for that day. (b) Prediction phase (visualized with the samples from the Samples + Prediction trial in (a)). Children predicted the apple quantities for the remaining 10 days, adding their guess for a day by repeatedly tapping the blue bar above that day, with one apple added for each tap. (c) Feedback phase. The true growth pattern (solid red apples) was superimposed on children's predictions (faint red apples), and the over-predicted apples were crossed out.

Test conditions

This study used a two (trial type: Prediction Only, and Sample + Prediction) by three (function family: linear, Gaussian, and exponential) within-subject design, with six trials in total. Each function family appeared in both trial types with different parameterizations. Children completed the Prediction Only trials before the Sample + Prediction trials. We chose this fixed order so that children would have similar priors about the possible function families before choosing their own data, as this knowledge could affect children's sampling.

Participants

Fifty 6- to 9-year-olds ($M_{\text{age}} = 7.84$, range = 5.75–9.75 years) were recruited from a local children's museum and a local school. Four participants were excluded from analyses due to data recording error ($n = 2$), experimenter error ($n = 1$), or task comprehension issue ($n = 1$), leaving 46 participants in the final sample. Due to the exploratory nature of this experiment and the constraints of testing in a museum or school setting, not all children completed the full task. Twenty-two children completed both trial types, 22 children only completed Prediction Only trials, and two children only completed Sample + Prediction trials ($M_{\text{trials}} = 3.74$, range = 2–6 trials).

Procedure

Children were tested in person, and the task was presented on a tablet device through a web-based interactive program. In the introduction phase, children learned about a farmer and his apple orchard by viewing a series of pictures. Children were told "Every day, the farmer picks all of the apples that grew in his orchard that day. He also keeps a record of how many apples grew each day. On some days, the weather is nice and a lot of apples grow. On other days, the weather is bad and only a few apples grow. So the number of apples that grow each day can be the same or different. Your job is to figure out the pattern of how apples grow from day to day."

The sampling phase was the only phase in which the two trial types differed (Figure 1a). In both trials, children first saw a timeline consisting of daily calendar icons labeled with numbers one through 15, representing 15 days. In Prediction Only trials, apple quantities for five of the 15 days were shown (all at once) through corresponding numbers of apples. Specifically, we chose days 2, 5, 8, 11, and 14 as a set of equidistant, informative samples. We confirmed that these samples were highly informative, through simulations which showed that they led to more accurate inferences than myopically optimal[1] or random samples. Children were told "The farmer shared part of his record with us, so we already know how many apples grew on these five days." In Sample + Prediction trials, no information about apple quantity was known initially. Children were told "The farmer has not shared his record with us, so you first need to choose five of these days to ask him about. When you ask him about a day, he will tell

[1] Samples chosen serially, with adaptation to newly sampled data; at each time step, the sample that maximally reduces uncertainty for the immediate next step (conditioning on all existing data) is chosen.

you how many apples grew that day. You can choose any five days to ask about." Children could sample a day by tapping above the corresponding calendar icon, and the apple quantity would appear and stay on the screen. Children repeated this process until they had sampled five different days.

In the prediction (test) phase, children guessed the apple quantities for the remaining 10 days (Figure 1b). Children made a prediction for a day by repeatedly tapping above the corresponding calendar icon—each tap would add one more apple. To distinguish the true apple quantities revealed during sampling and the predicted quantities, the former appeared as solid red apples while the latter appeared as faint red apples.

In the feedback phase, the true apple quantity for each of the 10 predicted days was superimposed on children's prediction as solid red apples (see Figure 1c). Over-predicted apples were crossed out, and under-predicted ones were circled. Children received qualitative feedback on their performance (e.g., "Some/most of your guesses were correct/incorrect.") so that they could focus on their overall performance instead of the exact prediction error at a given time point. Children were also asked to verbally describe the true apple growth pattern to ensure they understood the feedback.

Results

We first tested if children's prediction performance differed when they learned from passively received informative data (Prediction Only trials) versus self-selected data (Sample + Prediction trials), and across different function families. To measure children's performance on the prediction task, we computed the absolute prediction error for each of the 10 days that children made predictions for in a trial, by taking the absolute difference between the predicted and true apple quantities. Because this variable followed a Poisson distribution and each child made multiple predictions, we used generalized linear mixed models with a Poisson distribution, with a random intercept for each child to analyze their performance.

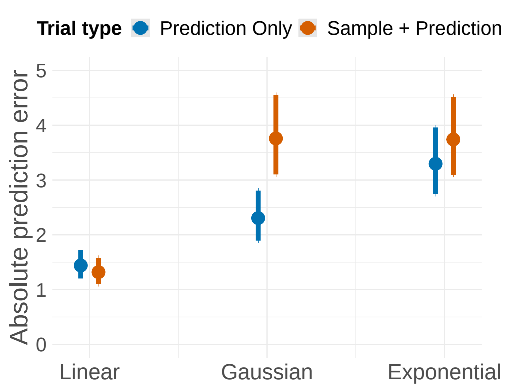

Figure 2: Absolute prediction error by function family. Vertical bars show 95% CIs.

We modeled children's absolute prediction error using trial type, function family, age (in months), and the interaction between trial type and function family as predictors (Figure 2). Children's prediction error varied significantly across trial types ($\chi^2(1) = 39.03$, $p < .001$) and function families ($\chi^2(2) = 391.74$, $p < .001$). Specifically, prediction error was lower in Prediction Only trials than Sample + Prediction trials ($b = -0.176$, 95% CI = $[-0.261, -0.09]$, $z = -4$, $p < .001$). Error was lower on the linear function than the Gaussian function ($b = -0.76$, 95% CI = $[-0.86, -0.66]$, $z = -14.6$, $p < .001$) and lower on the Gaussian function than the exponential function ($b = -0.18$, 95% CI = $[-0.24, -0.11]$, $z = -5.1$, $p < .001$). Children's prediction error also marginally decreased with age ($\chi^2(1) = 3.11$, $p = .078$). Further, the effect of trial type differed across function families ($\chi^2(2) = 43.32$, $p < .001$).

To investigate how children's learning of the three function families was differently affected by whether the data were passively received or self-selected, we conducted follow-up comparisons. Children had lower error in the Prediction Only trials than the Sample + Prediction trials when learning the Gaussian function ($b = -0.49$, 95% CI = $[-0.60, -0.38]$, $z = -8.86$, $p < .001$) and exponential function ($b = -0.13$, 95% CI = $[-0.23, -0.03]$, $z = -2.43$, $p = .015$). However, for the linear function, prediction error did not differ in the two types of trials ($b = -0.09$, 95% CI = $[-0.10, 0.27]$, $z = 0.94$, $p = .35$). In other words, the highly informative samples from the experimenter only led to more accurate inferences than the self-selected samples for the two non-linear functions, perhaps because children's error on the linear function was already low overall.

We then focused on the Sample + Prediction trials. We wanted to characterize children's sampling behavior, specifically how clustered or dispersed their samples were, to test if samples that spread across the function domain would provide a more accurate view of the function than dense samples that only cover a small region of the domain. To do so, we computed the standard deviation (SD) of children's five samples in each trial (e.g., if a child sampled days 2, 5, 8, 11, 14, the sample SD would be computed based on these numbers). The domain ranged from 1 to 15, so possible sample SD values ranged from 1.58 (sampling five immediately adjacent points) to 6.89 (sampling the two ends, e.g., days 1, 2, 3, 14, 15). A smaller sample SD would reflect sampling ad-

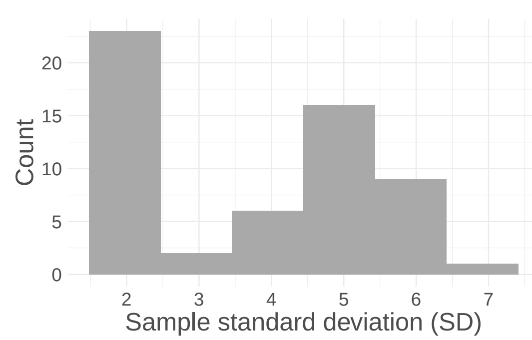

Figure 3: Distribution of trial-level sample SD.

jacent time points, whereas a larger value would reflect sampling more dispersed points. Children's sample SD followed a bimodal distribution (Figure 3). The major mode was between 1 and 2, roughly corresponding to a dense sampling strategy, while the minor mode was between 4 and 5, roughly corresponding to a dispersed sampling strategy (e.g., days 2, 5, 8, 11, 14 would have a sample SD of 4.74). We also characterized children's sampling behavior by clustering their samples, and the analysis revealed two clusters that corresponded to the two strategies above (see https://osf.io/h48m5/ for the clustering analysis). In theory, the sample SD measure by itself does not uniquely identify different strategies. However, in practice, it provides results that converge with our clustering analysis and captures the dispersion of the samples with an easily interpretable statistic. Therefore, we used sample SD as a measure of children's sampling strategy.

To test if the dispersed sampling strategy was more effective than the clustered strategy and if this effect varied across function families, we modeled absolute prediction error with sample SD, function family, age (in months), and the interaction between sample SD and function family as predictors. Prediction error decreased with a larger sample SD, or more dispersed samples ($\chi^2(1) = 63.35,\ p < .001$). Error also varied across function families ($\chi^2(2) = 124.07,\ p < .001$). Specifically, prediction error was lower for the linear function than both the Gaussian function ($b = -1.05,\ 95\%$ CI $= [-1.25, -0.86]$, $z = -10.5,\ p < .001$) and the exponential function ($b = -1.1,\ 95\%$ CI $= [-1.3, -0.9]$, $z = -10.8,\ p < .001$), but it did not differ for the latter two functions ($b = -0.05,\ 95\%$ CI $= [-0.16, 0.07]$, $z = 0.8,\ p = .42$). The effect of age did not reach significance ($\chi^2(1) = 1.19,\ p = .28$). Further, the effect of sample SD varied across function families ($\chi^2(2) = 14.76,\ p < .001$).

Children might have struggled with the non-linear functions partly because the clustered sampling strategy would be particularly misleading for inferring these functions. For example, sampling the first or last five points of the Gaussian function might lead a child to infer a roughly linear function

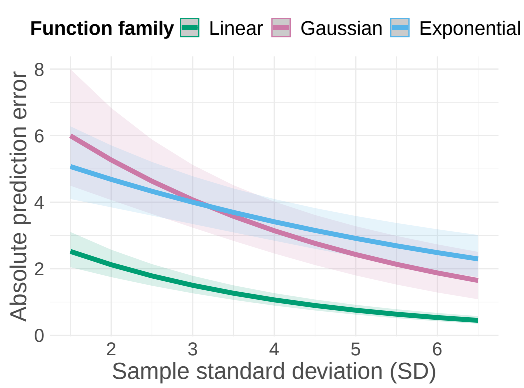

Figure 4: Absolute prediction error as a function of sample SD, with a separate Poisson regression for each function family. Colored bands show 95% CIs for the regressions.

(Figure 1a, right), while sampling dispersed points would reveal the non-monotonic nature of the function (Figure 1a, left). Thus, to test how sampling strategies might have affected children's predictions across function families, we analyzed the relationship between prediction error and sample SD separately for each function family (Figure 4). Unsurprisingly, given that linear functions can be inferred from any two points, sample SD did not affect prediction performance on the linear function ($b = -0.54,\ 95\%$ CI $= [-1.8, 0.5]$, $z = -1.13,\ p = .26$). However, a larger sample SD led to lower prediction error on the Gaussian function ($b = -0.25,\ 95\%$ CI $= [-0.38, -0.14]$, $z = -4.35,\ p < .001$) and exponential function ($b = -0.17,\ 95\%$ CI $= [-0.3, -0.045]$, $z = -2.76,\ p = .006$). Compared to clustered samples that only covered a small region of the domain, more dispersed samples might have helped children detect the defining features of the Gaussian function (i.e., non-monotonicity) and the exponential function (i.e., non-constant rates of change).

Lastly, we tested if children learned the correct abstract forms of the functions (e.g., the inverted U shape of a Gaussian function), regardless of if they learned the exact parameterizations (e.g., the specific height or location of the peak of a Gaussian function). To test if the function that a child predicted in a trial was best fit by the shape of the true function, we identified the best-fitting function from each of the three function families to the predicted function; we did so by minimizing the sum of squared error (SSE) between the candidate and predicted functions, using a BFGS optimizer (Nash, 2018). Out of the three candidate functions (one from each family), the one with the lowest SSE would be the one that best reflected the functional form that the child predicted.

For both the linear and Gaussian functions, most children's predictions were best fit by the shape of the true functions in both the Prediction Only and Sample + Prediction trials (Table 1). However, children's predictions for the exponential function were best fit by the linear function in Prediction Only trials, and all children who mislearned this function (across both trial types) represented a linear function instead. Nonetheless, 90% of the best-fitting functions that were classified as exponential (across both trial types) corresponded to true functions that were actually exponential, suggesting that a subset of children might have accurately distinguished the exponential function from the other functional forms. While children showed a bias towards inferring linear functions, under at least some circumstances, they correctly determined that functions were non-monotonic or non-linear.

Discussion

In an initial study, we showed that 6- to 9-year-old children can learn different families of functions using both highly informative data provided by an experimenter and self-selected data. In aggregate, children can distinguish between linear and non-linear functions and between monotonic and non-monotonic functions, by predicting distinct abstract forms for these functions. Children also learn linear functions bet-

(a) Prediction Only

True function		Best-fitting function: Linear	Gaussian	Exponential
	Linear	97.7%	–	2.3%
	Gaussian	28.1%	68.8%	3.1%
	Exponential	77.8%	–	22.2%

(b) Sample + Prediction

True function		Best-fitting function: Linear	Gaussian	Exponential
	Linear	100%	–	–
	Gaussian	45.5%	54.5%	–
	Exponential	45%	–	55%

Table 1: Confusion matrix for the functional form that best fit the predicted function for each function family, by trial type.

ter than non-linear ones, consistent with the linear bias that adults show in function learning (Brehmer, 1974; McDaniel & Busemeyer, 2005). Some children already show knowledge of the distinct features of non-monotonic and non-linear functions by predicting the correct shapes of the Gaussian and exponential functions. These results suggest that differing findings on children's understanding of non-linear or non-monotonic functions (e.g., Ebersbach & Resing, 2008; Ebersbach et al., 2008) may reflect individual differences or the sensitivity of children's ability to the task context.

The finding that some children select useful data (e.g., dispersed samples) and can rely on these data to draw more accurate inferences about the functions than children who select less useful data (e.g., clustered samples) is another testament to children's powerful active learning ability, which has been shown in many other contexts (e.g., Legare, 2012; L. E. Schulz & Bonawitz, 2007). This knowledge of sampling strategies also shares similarities with adults' intuition for choosing evenly-spaced samples, a generally helpful strategy for learning common functions (Gelpi et al., 2021, 2023). Despite the fact that children overall learn more accurately with passively received informative data than self-selected data, the high performance of this group of effective information seekers suggests that the effect of active learning may partly depend on the quality of the data that learners can gather.

Why might a significant minority of children select comparatively uninformative samples (i.e., in 37% of sampling trials, children selected immediately adjacent samples)? One possibility is that children are still learning to both predict and evaluate the usefulness of their samples. While clustered samples are not maximally informative, they are also not completely uninformative, especially early in the sampling process. Children may recognize that they are receiving new information, while failing to recognize that they could have learned even more. They may also struggle to compare possible sampling locations, in order to evaluate their relative informativeness. The small sample size of this initial study does not allow us to test how sampling strategies shift across trials or function families. In future studies, we plan to test if children flexibly adjust their sampling strategy based on its effectiveness in previous trials, as well as if this adjustment leads to better performance in subsequent trials.

Another possibility is that children who select uninformative samples may still have some nascent understanding of the efficacy of different sampling strategies. Children can identify efficient information-seeking questions (Ruggeri et al., 2017) before they can generate such questions (Ruggeri et al., 2016). Similarly, children can distinguish between confounded and unconfounded interventions when presented with contrasting choices (Lapidow & Walker, 2020; Sodian et al., 1991) before they can perform unconfounded interventions (Klahr et al., 2011; McCormack et al., 2016; Meng et al., 2018). In the future, by testing children's ability to choose between more or less informative samples (e.g., in a forced choice context), we could determine if children who do not yet implement effective sampling strategies may nevertheless recognize the usefulness of different strategies.

One concern about our task might be that children who select dispersed samples in the Sample + Prediction trials are simply copying the experimenter's sampling strategy in the Prediction Only trials. However, the fact that many children do not copy this strategy suggests that the ability to recognize the helpful strategy as worth copying may itself reflect knowledge of search strategies. Another concern might be that the response modality of the prediction task is action-intensive, as predicting apple quantities requires repeated tapping. Children may have been incentivized to underestimate the true values because less tapping is easier. However, the raw prediction error (signed) was in fact *above* 0 ($b = 0.689$, $t(51) = 3.87$, $p < .001$); children were generally engaged in the task and did not find the tapping onerous.

In future work, we also hope to study the developmental differences in children's and adults' active function learning. Learners' search behavior becomes less random across development, with early exploratory sampling followed by a narrower search (Blanco & Sloutsky, 2020; Gopnik et al., 2017; Hart et al., 2022; Lucas et al., 2014). When learning common functions, random sampling (e.g., of areas unlikely to provide additional information) may lead to wasted time and effort. However, when the true function to be learned is highly unusual, random sampling may lead to the accidental discovery of unexpected features of the environment. Development also affects learners' generalization ability, or how they make inferences based on existing information (Giron et al., 2023; E. Schulz et al., 2019). Thus, even children who collect adult-like samples that spread across the domain may draw different conclusions about the underlying function than adults.

Acknowledgments

We thank members of the Computational Cognitive Development Lab for their valuable feedback on this study. We also thank Anwyn Gatesy-Davis and Claire Washington for their help with recruitment and data collection. Finally, we thank the participating research sites and families.

References

Blanco, N. J., & Sloutsky, V. M. (2020). Attentional mechanisms drive systematic exploration in young children. *Cognition*, *202*, 104327.

Blanton, M., Brizuela, B. M., Gardiner, A. M., Sawrey, K., & Newman-Owens, A. (2015). A learning trajectory in 6-year-olds' thinking about generalizing functional relationships. *Journal for Research in Mathematics Education*, *46*(5), 511–558.

Blanton, M., & Kaput, J. J. (2004). Elementary grades students' capacity for functional thinking. *International Group For The Psychology Of Mathematics Education*.

Bott, L., & Heit, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 38.

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.

Ciccione, L., Sablé-Meyer, M., & Dehaene, S. (2022). Analyzing the misperception of exponential growth in graphs. *Cognition*, *225*, 105112.

Coates, N., Siegel, M., Tenenbaum, J., & Schulz, L. (2023). Representations of abstract relations in early childhood. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968.

Ebersbach, M., Lehner, M., Resing, W. C., & Wilkening, F. (2008). Forecasting exponential growth and exponential decline: Similarities and differences. *Acta Psychologica*, *127*(2), 247–257.

Ebersbach, M., & Resing, W. C. (2008). Implicit and explicit knowledge of linear and exponential growth in 5-and 9-year-olds. *Journal of Cognition and Development*, *9*(3), 286–309.

Ebersbach, M., Van Dooren, W., Goudriaan, M. N., & Verschaffel, L. (2010). Discriminating non-linearity from linearity: Its cognitive foundations in five-year-olds. *Mathematical Thinking and Learning*, *12*(1), 4–19.

Ebersbach, M., & Wilkening, F. (2007). Children's intuitive mathematics: The development of knowledge about nonlinear growth. *Child Development*, *78*(1), 296–308.

Gelpi, R., Saxena, N., Lifchits, G., Buchsbaum, D., & Lucas, C. G. (2021). Sampling heuristics for active function learning. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.

Gelpi, R., Zhou, C., Lucas, C., & Buchsbaum, D. (2023). Characterizing shifts in strategy in active function learning. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*.

Giron, A. P., Ciranka, S., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2023). Developmental changes in exploration resemble stochastic optimization. *Nature Human Behaviour*, *7*, 1955–1967.

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., Aboody, R., Fung, H., & 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*(30), 7892–7899.

Hart, Y., Kosoy, E., Liquin, E. G., Leonard, J. A., Mackey, A. P., & Gopnik, A. (2022). The development of creative search strategies. *Cognition*, *225*, 105102.

Kalish, M. L. (2013). Learning and extrapolating a periodic function. *Memory & Cognition*, *41*(6), 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.

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975.

Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315.

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Development*, *83*(1), 173–185.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

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.

McCormack, T., Bramley, N., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22.

McDaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review*, *12*(1), 24–42.

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science*, *24*(4), e13095.

Meng, Y., Bramley, N., & Xu, F. (2018). Children's causal interventions combine discrimination and confirmation. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Nash, J. C. (2018). *Compact numerical methods for computers: Linear algebra and function minimisation*. Routledge.

Ruggeri, A., Lombrozo, T., Griffiths, T. L., & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology*, *52*(12), 2159.

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). "Why is Toma late to school again?" Preschoolers identify the most informative questions. *Developmental Psychology*, *53*(9), 1620.

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive Psychology*, *99*, 44–79.

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.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045.

Shtulman, A., & Walker, C. (2020). Developing an understanding of science. *Annual Review of Developmental Psychology*, *2*, 111–132.

Sim, Z. L., & Xu, F. (2017). Learning higher-order generalizations through free play: Evidence from 2-and 3-year-old children. *Developmental Psychology*, *53*(4), 642.

Sobel, D. M., & Sommerville, J. A. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Psychology*, *1*, 176.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, *62*(4), 753–766.

Stephens, A. C., Fonger, N., Strachota, S., Isler, I., Blanton, M., Knuth, E., & Murphy Gardiner, A. (2017). A learning progression for elementary students' functional thinking. *Mathematical Thinking and Learning*, *19*(3), 143–166.

Weisberg, D. S., & Sobel, D. M. (2022). *Constructing science: Connecting causal reasoning to scientific thinking in young children*. MIT Press.

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

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*(12), 915–924.