Research Article

Psychological Science
2019, Vol. 30(11) 1561–1572

DOI: 10.1177/0956797619863663
www.psychologicalscience.org/PS

Searching for Rewards Like a Child Means Less Generalization and More Directed Exploration

Eric Schulz[1], Charley M. Wu[2], Azzurra Ruggeri[3,4], and Björn Meder[2,3,5]

[1]Department of Psychology, Harvard University; [2]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany; [3]Max Planck Research Group iSearch, Max Planck Institute for Human Development, Berlin, Germany; [4]School of Education, Technical University Munich; and [5]Department of Psychology, University of Erfurt

Abstract

How do children and adults differ in their search for rewards? We considered three different hypotheses that attribute developmental differences to (a) children's increased random sampling, (b) more directed exploration toward uncertain options, or (c) narrower generalization. Using a search task in which noisy rewards were spatially correlated on a grid, we compared the ability of 55 younger children (ages 7 and 8 years), 55 older children (ages 9–11 years), and 50 adults (ages 19–55 years) to successfully generalize about unobserved outcomes and balance the exploration–exploitation dilemma. Our results show that children explore more eagerly than adults but obtain lower rewards. We built a predictive model of search to disentangle the unique contributions of the three hypotheses of developmental differences and found robust and recoverable parameter estimates indicating that children generalize less and rely on directed exploration more than adults. We did not, however, find reliable differences in terms of random sampling.

Keywords

exploration–exploitation, development, generalization, search, multiarmed-bandit task, open data, open materials

Received 9/8/18; Revision accepted 6/21/19

Alan Turing (1950) famously believed that in order to build a general artificial intelligence, one must create a machine that can learn like a child. Indeed, recent advances in machine learning often contain references to childlike learning and exploration (Riedmiller et al., 2018). Yet little is known about how children actually explore and search for rewards in their environments and in what ways their behavior differs from that of adults.

In the course of learning through interactions with the environment, all organisms (biological or machine) are confronted with the *exploration–exploitation dilemma* (Mehlhorn et al., 2015). This dilemma highlights two opposing goals. The first goal is to explore unfamiliar options that provide useful information for future decisions yet may result in poor immediate rewards. The second goal is to exploit options known to have high expectations of reward but potentially forgo learning about unexplored options.

In addition to balancing exploration and exploitation, another crucial ingredient for adaptive search behavior is a mechanism that can generalize beyond observed outcomes, thereby guiding search and decision making by forming inductive beliefs about novel options. For example, from a purely combinatorial perspective, it takes only a few features and a small range of values to generate a pool of options vastly exceeding what could ever be explored in a lifetime. Nonetheless, humans of all ages manage to generalize from limited experiences in order to choose from among a set of potentially unlimited possibilities. Thus, a model of human search also needs to provide a mechanism for generalization.

Previous research has found extensive variability and developmental differences in children's and adults' search behavior, which not only result from a progressive refinement of basic cognitive functions (e.g., memory,

Corresponding Author:
Eric Schulz, Harvard University, Department of Psychology, 52 Oxford St., Cambridge, MA 02138
E-mail: ericschulz@fas.harvard.edu

attention) but also derive from systematic changes in the computational principles driving behavior (Palminteri, Kilford, Coricelli, & Blakemore, 2016). In particular, developmental differences in learning and decision making have been explained by appealing to three hypothesized mechanisms: Children sample more randomly, explore more eagerly, and generalize more narrowly than adults.

In this study, we investigated how these three mechanisms are able to explain developmental differences in exploration–exploitation behavior. We provided a precise characterization of these competing ideas in a formal model, which was used to predict behavior in a search task in which noisy and continuous rewards were spatially correlated. Using behavioral markers, interpreting parameter estimates from computational models, and analyzing judgments about unexplored options, we found that children generalize less but engage in more directed exploration than adults. We did not, however, find reliable developmental differences in random exploration. These results enrich our understanding of maturation in learning and decision making, demonstrating that children explore using uncertainty-guided mechanisms rather than simply behaving more randomly.

A Tale of Three Mechanisms

Development as cooling off

Because optimal solutions to the exploration–exploitation dilemma are generally intractable (Bellman, 1952), heuristic alternatives are frequently employed. In particular, learning under the demands of the exploration–exploitation trade-off has been described using at least two distinct strategies (Wilson, Geana, White, Ludvig, & Cohen, 2014). One such strategy is increased *random exploration*, which uses noisy, random sampling to learn about new options.

A key finding in the psychological literature is that children tend to try out more options than adults (Cauffman et al., 2010; Mata, Wilke, & Czienskowski, 2013). This has been interpreted as evidence for higher levels of random exploration in children and has been loosely compared with algorithms of simulated annealing from computer science (Gopnik et al., 2017), in which the amount of random exploration gradually reduces over time. Children can be described as having higher temperature parameters, in which the learner initially samples very randomly across a large set of possibilities before eventually focusing on a smaller subset (Gopnik, Griffiths, & Lucas, 2015). This temperature parameter is expected to "cool off" with age, leading to lower levels of random exploration in late childhood and adulthood.

Development as reduction of directed exploration

A second strategy to tackle the exploration–exploitation dilemma is to use *directed exploration* by preferentially sampling highly uncertain options in order to gain more information and reduce uncertainty about the environment. Directed exploration has been formalized by introducing an "uncertainty bonus" that values the exploration of lesser known options (Auer, 2002), with behavioral markers found in a number of studies (Frank, Doll, Oas-Terpstra, & Moreno, 2009; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018).

Directed exploration treats information as intrinsically valuable by inflating rewards by their estimated uncertainty (Auer, 2002). This leads to a more sophisticated *uncertainty-guided sampling* strategy that could also explain developmental differences. Indeed, the literature on self-directed learning shows that children are clearly capable of exploring their environment in a systematic, directed fashion. Already, infants tend to value the exploration of uncertain options (L. Schulz, 2015), and children can balance theory and evidence in simple exploration tasks (Bonawitz, van Schijndel, Friel, & Schulz, 2012) and are able to efficiently adapt their search behavior to different environmental structures (Ruggeri & Lombrozo, 2015). Moreover, children can sometimes even outperform adults in the self-directed learning of unusual relationships (Lucas, Bridgers, Griffiths, & Gopnik, 2014). Both directed and random exploration do not have to be mutually exclusive mechanisms, with recent research finding signatures of both types of exploration in adolescent and adult participants (Gershman, 2018; Somerville et al., 2017; Wilson et al., 2014).

Development as refined generalization

Rather than explaining development as a change in how we explore given some beliefs about the world, *generalization-based accounts* attribute developmental differences to the way we form our beliefs in the first place. Many studies have shown that human learners use structured knowledge about the environment to guide exploration (E. Schulz, Konstantinidis, & Speekenbrink, 2017), where the quality of these representations and the way that people use them to generalize across experiences can have a crucial impact on search behavior. Thus, development of more complex cognitive processes (Blanco et al., 2016), leading to broader generalizations, could also account for the observed developmental differences in sampling behavior.

The notion of generalization as a mechanism for explaining developmental differences has a long-standing history in psychology. For instance, Piaget (1964) assumed

that children learn and adapt to different situational demands by the processes of assimilation (applying a previous concept to a new task) and accommodation (changing a previous concept in the face of new information). Expanding on Piaget's idea, Klahr (1982) proposed generalization as a crucial developmental process, in particular the mechanism of regularity detection, which supports generalization and improves over the course of development. More generally, the implementation of various forms of decision making (Hartley & Somerville, 2015) could be constrained by the capacity for complex cognitive processes, which become more refined over the life span. For example, although younger children attend more frequently to irrelevant information than older children (Hagen & Hale, 1973), they can be prompted to attend to the relevant information by marking the most relevant cues, whereupon they eventually select the best alternative (Davidson, 1996). Thus, children may indeed be able to apply uncertainty-driven exploratory strategies but lack the appropriate task representation to successfully implement them.

A Task to Study Generalization and Exploration

We studied the behavior of both children and adults in a spatially correlated multiarmed-bandit task (Wu et al., 2018; see Fig. 1a), in which rewards were distributed on a grid characterized by spatial correlation (i.e., similar rewards cluster together; see Fig. 1g; for a similar task, see White, 2013), and the search horizon was vastly smaller than the number of options. Efficient search and accumulation of rewards in such an environment require two critical components. First, participants need to learn about the underlying spatial correlation in order to generalize from observed rewards to unseen options. This is crucial because there are considerably more options than can be explored within the limited search horizon. Second, participants need a sampling strategy that achieves a balance between exploring new options and exploiting known options with high rewards.

Method

Participants

We recruited 55 younger children (26 female; age: M = 7.53 years, SD = 0.50, range = 7–8), 55 older children (24 female; age: M = 9.95 years, SD = 0.80, range = 9–11), and 50 adults (25 female; age: M = 33.76 years, SD = 8.53, range = 18–55) at the Berlin Natural History Museum in Germany. We determined the different age groups and the number of participants per group before data collection on the basis of existing findings showing strong developmental differences between ages 7 and 10 years in children's question asking and active search behavior (Davidson, 1991; Ruggeri & Lombrozo, 2015). Participants were paid up to €3.50 for taking part in the experiment, contingent on performance (M = €2.67, SD = 0.50, range = €2.00–€3.50). Informed consent was obtained from all participants.

Design

The experiment used a between-subjects design, in which each participant was randomly assigned to one of two different classes of environments (see Fig. 1g): smooth or rough, with smooth environments having stronger spatial correlations than rough environments. We generated 40 of each class of environments from a radial-basis-function kernel (see below), with λ of 4 for smooth and λ of 1 for rough. On each round, a new environment was sampled (without replacement) from the set of 40 environments, which was then used to define a bivariate function on the grid, with each observation including additional normally distributed noise $\epsilon \sim \mathcal{N}(0, 1)$. The task was presented over 10 rounds on different grid worlds drawn from the same class of environments. The first round was a tutorial round, and the last round was a bonus round, in which participants sampled for 15 trials and then had to generate predictions for five randomly chosen and previously unobserved tiles on the grid. Participants had a search horizon of 25 trials per grid, including repeat clicks.

Materials and procedure

Participants were introduced to the task through a tutorial round, which familiarized them with the spatial correlation of rewards and the possibility of reclicking tiles. Moreover, participants were told that they would be rewarded on the basis of the sum of sampled points. Afterward, they had to complete three comprehension questions before starting the task. At the beginning of each round, one random tile was revealed, and participants could click on any of the tiles (including reclicks) on the grid until the search horizon was exhausted. Clicking an unrevealed tile displayed the numerical value of the reward along with a corresponding color aid; darker colors indicated higher rewards. Per round, observations were scaled to a randomly drawn maximum value in the range of 35 to 45 so that the value of the global optima could not be easily guessed. Reclicked tiles could show some variations in the observed value because of noise. For repeat clicks, the most recent observation was displayed numerically, and the color of the tile corresponded to the mean of all

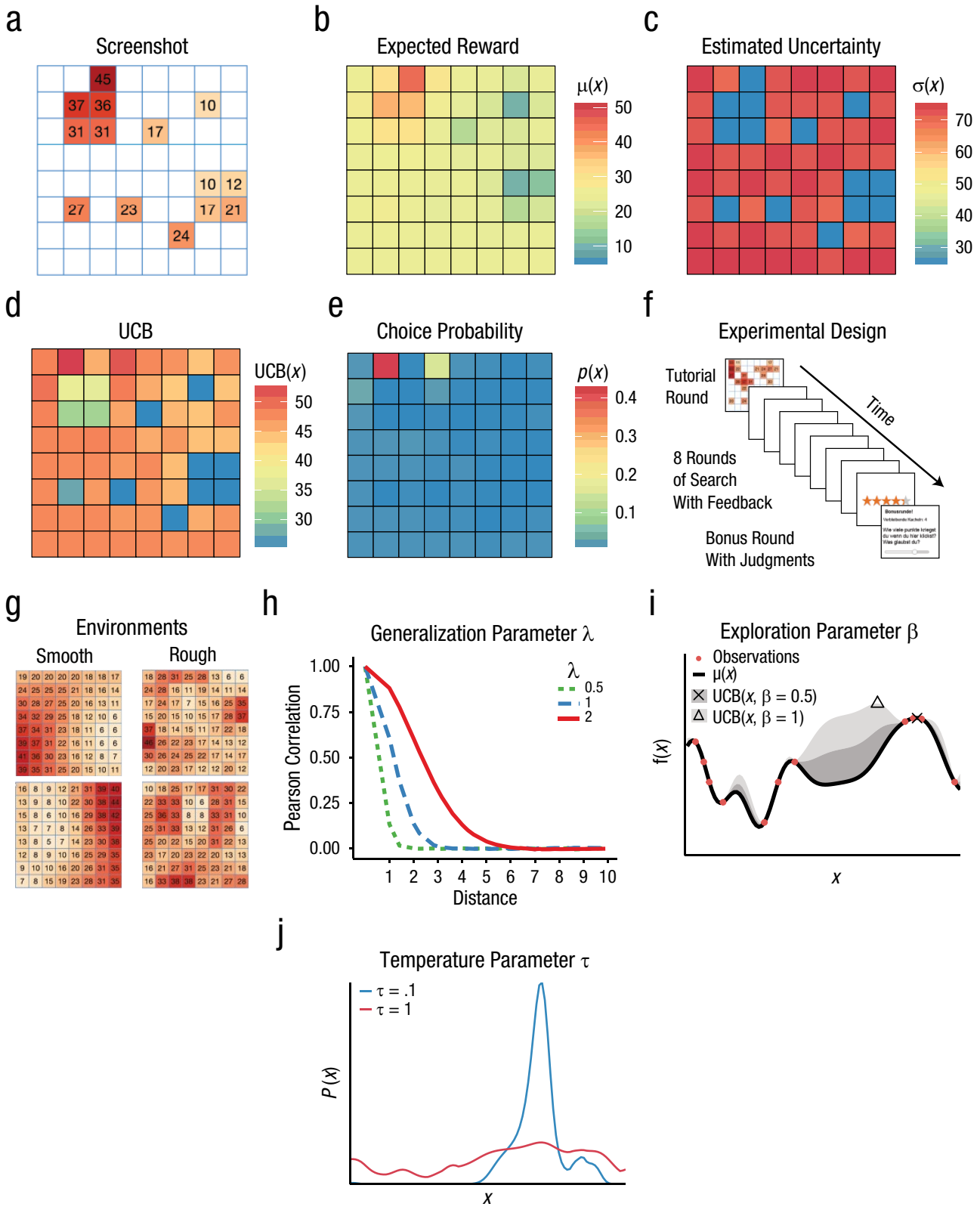

Fig. 1. *(continued on next page)*

Fig. 1. Overview of task and model. The screenshot (a) shows the experiment in the middle of a round with the grid partially revealed. Expected reward (b) and estimated uncertainty (c) based on observations of the grid in (a) are shown as obtained using Gaussian-process regression as a model of generalization. Upper confidence bounds (UCBs) for each option (d) are based on a weighted sum of (b) and (c). Choice probabilities of the *softmax* function are shown in (e). Median participant parameter estimates are used in (b) through (e). An overview of the experimental design is shown in (f). Participants first completed a tutorial round and then eight rounds of search with feedback. Finally, they completed a bonus round that also included judgments about unobserved tiles. The two types of environments used in the experiment are shown in (g): Smooth environments had stronger spatial correlations than rough environments. Correlations of rewards between different options (h) decay exponentially as a function of their distance, where higher values of λ lead to slower decays and broader generalizations. The illustration of UCB sampling (i) uses a univariate example, in which the expected reward (black line) and estimated uncertainty (gray ribbons for different values of β) are summed. Higher values of β value the exploration of uncertain options more strongly (compare the arguments of the maxima of the two β values, indicated by the × and the triangle). In the overview of the *softmax* function (j), higher values of the temperature parameter τ lead to greater random exploration.

previous observations. In the bonus round, participants sampled for 15 trials and were then asked to generate predictions for five randomly selected and previously unobserved tiles. This was explained to them before the bonus round started. Additionally, participants had to indicate how certain they were about each prediction on a scale from 0 to 10. Afterward, they had to select one of the five tiles before continuing with the round.

Participants were awarded up to 5 stars at the end of each round (e.g., 4.6 out of 5) on the basis of the ratio of their average reward to the global maximum. The performance bonus was calculated on the basis of the average number of stars earned in each round, excluding the tutorial round: 5 out of 5 stars corresponded to €3.50, whereas each half-star interval reduced the bonus by €0.50 until a minimum bonus of €0.50.

A combined model of generalization and exploration

We used a formal model that combined generalization with a sampling strategy accounting for both directed and random exploration (Wu et al., 2018) to predict each participant's out-of-sample search behavior. The generalization component was based on Gaussian-process regression, which is a Bayesian function-learning approach theoretically capable of learning any stationary function (Rasmussen & Williams, 2006) and has been found to effectively describe human behavior in explicit function-learning tasks (Lucas, Griffiths, Williams, & Kalish, 2015). The Gaussian-process component is used to adaptively learn a value function, which generalizes the limited set of observed rewards over the entire search space using Bayesian inference.

The Gaussian-process prior is completely determined by the choice of a kernel function, $k(x, x')$, which encodes assumptions about how points in the input space are related to each other. A common choice of this function is the *radial-basis function*:

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\lambda}\right),$$

where the length-scale parameter λ encodes the extent of spatial generalization between options (tiles) in the grid. The assumptions of this kernel function are similar to the gradient of generalization historically described by Shepard (1987), which also models generalization as an exponentially decaying function of the stimulus similarity distance (see Fig. 1h), which has been observed across a wide range of stimuli and organisms. As an example, generalization with λ of 1 corresponds to the assumption that the rewards of two neighboring tiles are correlated by an r of .6 and that this correlation effectively decays to 0 for options more than three tiles apart. We treated λ as a free parameter in our model comparison to assess age-related differences in the capacity for generalization.

Given different possible options (x) to sample from (i.e., tiles on the grid), Gaussian-process regression generated normally distributed beliefs about rewards with expectation $\mu(x)$ and estimated uncertainty $\sigma(x)$; see Figures 1b and 1c. A sampling strategy was then used to map the beliefs of the Gaussian process onto a valuation for sampling each option at a given time. Crucially, such a sampling strategy must address the exploration–exploitation dilemma. One frequently applied heuristic for solving this dilemma is upper-confidence-bound (UCB) sampling (Srinivas, Krause, Kakade, & Seeger, 2009), which evaluates each option on the basis of a weighted sum of expected reward and estimated uncertainty:

$$\mathrm{UCB}(x) = \mu(x) + \beta\sigma(x),$$

where β models the extent to which uncertainty (in addition to mean rewards) is valued positively and therefore directly sought out. This strategy corresponds to directed exploration because it encourages the sampling of options with higher uncertainty according to the underlying generalization model (see Fig. 1i). We treated the exploration parameter β as a free parameter to assess how much participants value the reduction of uncertainty (i.e., engage in directed exploration). As an example, an exploration bonus β of 0.5 means that participants would prefer option x_1, expected to have

reward $\mu(x_1)$ equal to 30 and uncertainty $\sigma(x_1)$ equal to 10, over option x_2, expected to have reward $\mu(x_2)$ equal to 34 and uncertainty $\sigma(x_2)$ equal to 1. This is because sampling x_1 is expected to reduce a larger amount of uncertainty, even though x_2 has a higher expected reward: $\mathrm{UCB}(x_1 \mid \beta = 0.5) = 35$ versus $\mathrm{UCB}(x_2 \mid \beta = 0.5) = 34.5$.

Finally, we use a *softmax* function to map the UCB values, UCB(x), of our proposed Gaussian-process–UCB sampling model onto choice probabilities:

$$p(x) = \frac{\exp\left(\mathrm{UCB}(x)/\tau\right)}{\sum_{j=1}^{N} \exp\left(\mathrm{UCB}(x_j)/\tau\right)},$$

where τ is the temperature parameter governing the amount of randomness in sampling behavior. If τ is high (higher temperatures), then participants are assumed to sample more randomly, whereas if τ is low (cooler temperatures), the choice probabilities are concentrated on the highest valued options (see Fig. 1j). Thus, τ encodes the tendency toward random exploration. We treated τ as a free parameter to assess the extent of random exploration in children and adults (for alternative implementations such as ϵ-greedy sampling and estimation of optimal parameters, see the Supplemental Material available online).

In summary, Gaussian-process–UCB models contain three different parameters: the length-scale λ capturing the extent of generalization, the exploration bonus β describing the extent of directed exploration, and the temperature parameter τ modulating random exploration. These three parameters directly correspond to the three postulated mechanisms of developmental differences in various decision-making tasks and can also be robustly recovered (see the Supplemental Material).

Results

Behavioral results

Comparing participants' average rewards, we found that participants gained higher rewards in smooth than in rough environments (see Fig. 2a), $t(158) = 10.51$, $p < .001$, $d = 1.66$, 95% confidence interval (CI) = [1.30, 2.02], Bayes factor (BF) > 100, suggesting that they made use of the spatial correlations and performed better when correlations were stronger. Adults performed better than older children (see Fig. 2a), $t(103) = 4.91$, $p < .001$, $d = 0.96$, 95% CI = [0.55, 1.37], BF > 100, who in turn performed somewhat better than younger children, $t(108) = 2.42$, $p = .02$, $d = 0.46$, 95% CI = [0.08, 0.84], BF = 2.68. Analyzing the distance between consecutive choices (see Fig. 2b) revealed that participants sampled more locally (smaller distances) in smooth than in rough environments, $t(158) = -3.83$, $p < .001$, $d = 0.61$, 95% CI = [0.29, 0.93], BF > 100. Adults sampled more locally than older children, $t(103) = -3.9$, $p < .001$, $d = 0.76$, 95% CI = [0.36, 1.16], BF > 100, but there was no difference between younger and older children, $t(108) = 1.76$, $p = .08$, $d = 0.34$, 95% CI = [−0.05, 0.72], BF = 0.80. Importantly, adults sampled fewer unique options than older children (14.5 vs. 21.7), $t(103) = -6.77$, $d = 1.32$, 95% CI = [0.90, 1.75], $p < .001$, BF > 100, whereas the two children groups did not differ in how many unique options they sampled (21.7 vs. 22.7), $t(108) = 1.27$, $d = 0.24$, 95% CI = [−0.14, 0.62], $p = .21$, BF = 0.4.

Looking at the learning curves (i.e., average rewards over trials; see Fig. 2c), we found a positive rank correlation between mean rewards and trial number, Spearman's $\rho = .12$, 95% CI = [.08, .16], $t(159) = 6.12$, $p < .001$, BF > 100. Although this correlation did not differ between the rough and smooth conditions, $t(158) = -0.43$, $p = .67$, $d = 0.07$, 95% CI = [−0.24, 0.38], BF = 0.19, it was significantly higher for adults than for older children (.29 vs. .08), $t(103) = 5.90$, $p < .001$, $d = 1.15$, 95% CI = [0.74, 1.57], BF = 0.19, BF > 100. The correlation between trials and rewards did not differ between younger and older children (.04 vs. .08), $t(108) = -1.87$, $p = .06$, $d = 0.36$, 95% CI = [−0.02, 0.74], BF = 0.96. Therefore, adults learned faster, whereas children explored more extensively (for further behavioral analyses, see the Supplemental Material).

Model comparison

We compared the Gaussian-process–UCB model with an alternative model that does not generalize across options but is a powerful Bayesian model for reinforcement learning across independent reward distributions (*mean-tracker* model). Model comparisons were based on leave-one-round-out cross-validation error, in which we fitted each model combined with the UCB sampling strategy to each participant using a training set omitting one round, and then we assessed predictive performance on the hold-out round. Repeating this procedure for every participant and all rounds (apart from the tutorial and the bonus rounds), we calculated the standardized predictive accuracy for each model (pseudo R^2 comparing out-of-sample log loss with random chance), where 0 indicates chance-level predictions, and 1 indicates theoretically perfect predictions (for full model comparison with additional sampling strategies, see the Supplemental Material). The results of this comparison are shown in Figure 2d. The Gaussian-process–UCB model predicted participants' behavior better overall, $t(159) = 13.28$, $p < .001$, $d = 1.05$, 95% CI = [0.82, 1.28], BF > 100, and also for adults, $t(49) = 5.98$, $p < .001$, $d = 0.85$, 95% CI = [0.43, 1.26], BF > 100; older children, $t(54) = 10.92$, $p < .001$, $d = 1.48$, 95% CI = [1.05, 1.90], BF > 100; and younger children, $t(54) =$

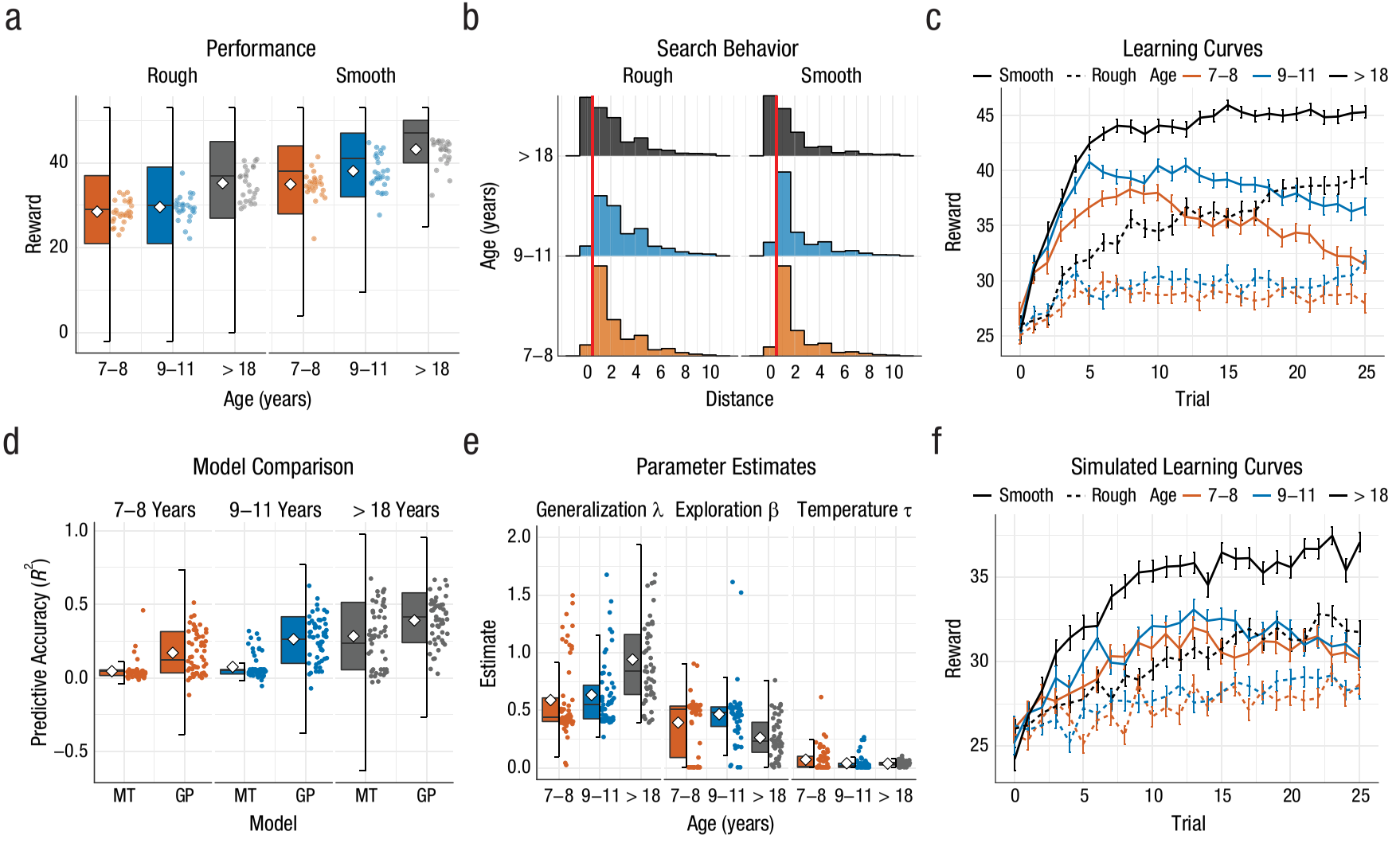

Fig. 2. Main results. Tukey box-and-whisker plots of rewards (a) show the distribution of all choices for all participants, separately for each age group and condition. In each box, the horizontal line represents the median, the height of the box shows the interquartile range of the distribution, and the whiskers show 1.5 times the interquartile range. Each dot is a participant-wise mean, and diamonds indicate group means. Histograms (b) show distances between consecutive choices by age group and condition. A distance of zero corresponds to a repeat click. The vertical red line marks the difference between a repeat click and sampling a different option. Mean reward across trials (c) is shown as a function of condition and age group. Error bars indicate standard errors of the mean. Tukey box-and-whisker plots for model comparisons (d) show predictive-accuracy distributions for Gaussian-process (GP) and mean-tracker (MT) models by age group. Tukey box-and-whisker plots of cross-validated parameters retrieved from the GP–upper-confidence-bound (UCB) model (e) are shown for each age group. In (d) and (e), each dot is a participant-wise mean, diamonds indicate group means, horizontal lines indicate medians, the height of each box shows the interquartile range of the distribution, and the whiskers show 1.5 times the interquartile range. Outliers have been removed for readability but were included in all statistical tests (see the Supplemental Material available online). Learning curves simulated by the GP-UCB model using mean participant parameter estimates (f) are shown as a function of trial, condition, and age group. Error bars indicate standard errors of the mean.

6.77, $p < .001$, $d = 0.91$, 95% CI = [0.52, 1.31], BF > 100. The Gaussian-process–UCB model predicted adults' behavior better than that of older children, $t(103) = 4.33$, $p < .001$, $d = 0.85$, 95% CI = [0.44, 1.25], BF > 100, which in turn was better predicted than behavior of younger children, $t(108) = 3.32$, $p = .001$, $d = 0.63$, 95% CI = [0.24, 1.02], BF = 24.8.

Developmental differences in parameter estimates

We analyzed the mean participant parameter estimates of the Gaussian-process–UCB model (see Fig. 2e) to assess the contributions of the three mechanisms (generalization, directed exploration, and random exploration) toward developmental differences. We found that adults generalized more than older children, as indicated by larger λ estimates, Mann-Whitney $U = 2{,}001$, $p < .001$, $r_\tau = .32$, 95% CI = [.18, .47], BF > 100, whereas the two groups of children did not differ significantly in their extent of generalization, $U = 1{,}829$, $p = .06$, $r_\tau = .15$, 95% CI = [−.01, .30], BF = 1.7. Furthermore, older children valued the reduction of uncertainty more than adults (i.e., higher β values), $U = 629$, $p < .001$, $r_\tau = .39$, 95% CI = [.25, .52], BF > 100, whereas there was no difference between younger and older children, $U = 1{,}403$, $p = .51$, $r_\tau = .05$, 95% CI = [−.10, .21], BF = 0.2. Critically, whereas there were strong differences between children and adults for the parameters capturing generalization and directed exploration, there was no reliable difference in the *softmax* temperature parameter τ, with no difference between older children and adults, $W = 1{,}718$, $p = .03$, $r_\tau = .17$, 95% CI = [.01, .34], BF = 0.7, and only anecdotal differences between the two groups of children, $W = 1{,}211$, $p = .07$, $r_\tau = .14$, 95% CI = [−.01, .30], BF = 1.4.[1] This suggests that the amount of random exploration did not reliably differ by age group (for other implementations of random exploration, see the Supplemental Material). Thus, our modeling results converged on the same conclusion as the behavioral results. Children explore more than adults, yet instead of being random, children's exploration behavior seems to be directed toward options with high uncertainty. Additionally, our parameter estimates are robustly recoverable (see the Supplemental Material) and can be used to simulate learning curves that reproduce the differences between the age groups as well as between smooth and rough conditions (see Fig. 2f).

Bonus round

In the bonus round, each participant predicted the expected rewards and the underlying uncertainty for five randomly sampled unrevealed tiles after having made 15 choices on the grid. We first calculated the mean absolute error between predictions and the true expected value of rewards (see Fig. 3a). Prediction error was higher for rough compared with smooth environments, $t(158) = 4.93$, $p < .001$, $d = 0.78$, 95% CI = [0.46, 1.10], BF > 100, reflecting the lower degree of spatial correlation that could be used to evaluate unseen options. Surprisingly, older children were as accurate as adults, $t(103) = 0.28$, $p = .78$, $d = 0.05$, 95% CI = [−0.44, 0.33], BF = 0.2, but younger children performed worse than older children, $t(108) = 3.14$, $p = .002$, $d = 0.60$, 95% CI = [0.21, 0.99], BF = 15. Certainty judgments did not differ between the smooth and rough environments, $t(158) = 1.13$, $p = .26$, $d = 0.18$, 95% CI = [−0.13, 0.49], BF = 0.2, or between the different age groups (maximum BF = 0.1).

Of particular interest is how judgments about the expectation of rewards and perceived uncertainty relate to the eventual choice from among the five options (implemented as a five-alternative forced choice). We standardized the estimated reward and confidence judgment of each participant's chosen tile by dividing by the sum of the estimates for all five options (see Fig. 3c). Thus, larger standardized estimates reflect a larger contribution of either high reward or high certainty on the choice. Whereas there was no difference between age groups in terms of the estimated reward of the chosen option (maximum BF = 0.1), we found that younger children preferred options with higher uncertainty slightly more than older children, $t(108) = 2.22$, $p = .03$, $d = 0.42$, 95% CI = [0.04, 0.80], BF = 1.8, and substantially more than adults, $t(103) = 2.82$, $p = .006$, $d = 0.55$, 95% CI = [0.16, 0.95], BF = 6.7. This further corroborates our previous analyses, showing that the sampling behavior of children is more directed toward uncertain options than that of adults.

Discussion

We examined three potential sources of developmental differences in a complex learning and decision-making task: random exploration, directed exploration, and generalization. Using a paradigm that combines both generalization and search, we found that adults gained higher rewards and exploited more strongly, whereas children sampled more unique options, thereby gaining lower rewards but exploring the environment more extensively. Using a computational model with parameters directly corresponding to the three hypothesized mechanisms of developmental differences, we found that children generalized less and were guided by directed exploration more strongly than adults. They did not, however, explore more randomly than adults.

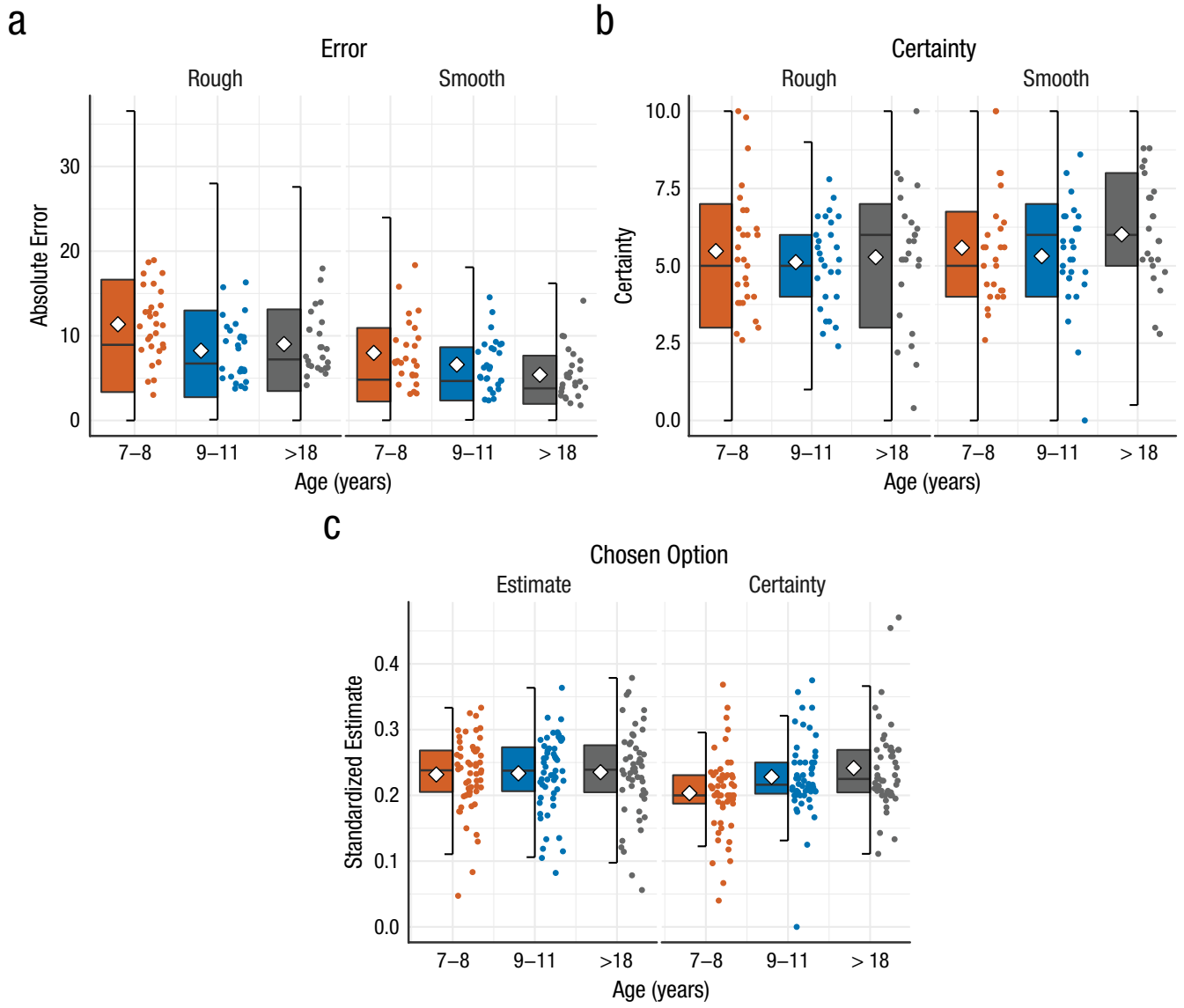

Fig. 3. Bonus-round results. Tukey box-and-whisker plots show (a) mean absolute error of participant predictions about the rewards of unobserved tiles, (b) certainty judgments, and (c) standardized predictions and certainty estimates as a function of age group. Results in (a) and (b) are further separated by age group, whereas estimates in (c) are separated by how much the estimated reward and certainty influenced choice (relative to judgments about nonchosen options). In (b), values on the *y*-axis run from least certain (0) to most certain (10). In all panels, dots are participant means, diamonds are group means, horizontal lines inside boxes represent medians, each box shows the interquartile range of the distribution, and the whiskers show 1.5 times the interquartile range.

Our results shed new light on the developmental trajectories in generalization and exploration, casting children not as merely prone to more random sampling behavior but as directed explorers who are hungry for information in their environment. Our conclusions are drawn from converging evidence combining analysis of behavioral data and computational modeling. Moreover, our findings are highly recoverable and also hold for other formalizations of random exploration instead of using the *softmax* temperature parameter (see the Supplemental Material).

Interestingly, related work by Somerville et al. (2017) also found no developmental difference in random exploration but increasing directed exploration across early adolescence, which stabilized in adulthood. We believe that our results are not necessarily incompatible with that finding. Somerville and colleagues defined directed exploration using horizon-sensitive exploration (i.e., strategic planning of exploration), whereas we defined directed exploration as uncertainty-guided exploration via a greedy UCB algorithm. Thus, children may have higher tendencies toward directed exploration

in a stepwise greedy fashion but fail to exhibit such tendencies when planning ahead for multiple steps, perhaps because of cognitive limitations. This opens up further possibilities for studying different mechanisms of directed exploration and how they relate to one another.

Our results provide strong evidence for developmental differences in directed exploration driven by both expected rewards and the associated uncertainty. These findings complement existing research on age-related differences in risk- and uncertainty-related behavior (Josef et al., 2016). For instance, adolescents and adults systematically differ in their tolerance of options with outcomes that have unknown probabilities, providing converging evidence that uncertainty is valued differently depending on age (Tymula et al., 2012). Importantly, in our task, a sampling strategy that sought only to reduce uncertainty was inferior to the "optimistic" UCB strategy in predicting children's and adults' behavior (for details, see the Supplemental Material). This result demonstrates how reward expectations and uncertainty interact to produce decision-making behavior that balances the exploration–exploitation trade-off adaptively as a function of age. Future work should attempt to further disentangle different interpretations of uncertainty seeking formally, for example, by not familiarizing participants with the underlying environments or by manipulating the level of noise in the outcomes directly.

Furthermore, it is surprising that there were no meaningful differences between younger and older children's parameter estimates. Because this indicates that directed exploration might be present even earlier than expected, future studies could apply our paradigm to investigate exploration behavior in even younger children.

Our results showing a developmental increase in generalization can also be related to previous findings showing a developmental increase in the use of task-structure knowledge in model-based reward learning (Decker, Otto, Daw, & Hartley, 2016). Because the generalization parameter λ can be mathematically equated to the speed of learning about the underlying function (Sollich, 1999), generalization and learning are inextricably linked in our task. There are, however, other uses of the term *generalization* in the psychological literature. For example, children are known to generalize words or categories more broadly, a tendency that decreases over time, trading off with the capacity to form more precise episodic memories (Keresztes et al., 2017). Whereas we focused on generalization in the sense used by Shepard (1987; i.e., generalization across stimuli), it is an outstanding question how this type of generalization relates to word and category generalization. It would be a fruitful avenue for future research to connect these two domains in a unifying theory of generalization.

In our current study, we assessed environments with only stationary reward distributions. However, given that children displayed increased exploration behavior, we believe that they could perform especially well in environments that change over rounds. Whether or not children would outperform adults in changing environments remains an important question for future research. Ultimately, our results suggest that to fulfill Alan Turing's dream of creating a childlike artificial intelligence, we need to incorporate generalization and curiosity-driven exploration mechanisms.

Action Editor

Erika E. Forbes served as action editor for this article.

Author Contributions

All the authors developed the study concept and contributed to the study design. E. Schulz and C. M. Wu analyzed and interpreted the data under the supervision of B. Meder and A. Ruggeri. E. Schulz and C. M. Wu drafted the manuscript, and B. Meder and A. Ruggeri provided critical revisions. All the authors approved the final manuscript for submission.

ORCID iDs

Eric Schulz https://orcid.org/0000-0003-3088-0371
Charley M. Wu https://orcid.org/0000-0002-2215-572X
Azzurra Ruggeri https://orcid.org/0000-0002-0839-1929

Acknowledgments

We thank all of the families who participated in this research; the Berlin Natural History Museum, where we conducted the study; Andreas Sommer for collecting the data; and Federico Meini for help with programming the experiment.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by the Max Planck Society and Deutsche Forschungsgemeinschaft Grant No. ME 3717/2-2 to B. Meder. E. Schulz is supported by the Harvard Data Science Initiative. C. M. Wu is supported by the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619863663

Open Practices

All data, analysis codes, and code needed to reproduce the experiment have been made publicly available and can be accessed at https://osf.io/3qskj/. The design and analysis plans for this study were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619863663. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Note

1. We also assessed whether there was a correlation between age and parameter estimates for the adult participants. This analysis revealed no relation between age and λ, $r = -.11$, $t(48) = -0.73$, $p = .47$, BF = 0.4; age and β, $r = .15$, $t(48) = -1.03$, $p = .31$, BF = 0.5; or age and τ, $r = -.09$, $t(48) = -0.62$, $p = .53$, BF = 0.4. However, these results should be interpreted with caution because they are based on data from only 50 participants. Future research should try to further map out the developmental trajectories of these parameters across the whole life span.

References

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, *3*, 397–422.

Bellman, R. (1952). On the theory of dynamic programming. *Proceedings of the National Academy of Sciences, USA*, *38*, 716–719.

Blanco, N. J., Love, B. C., Ramscar, M., Otto, A. R., Smayda, K., & Maddox, W. T. (2016). Exploratory decision-making as a function of lifelong experience, not cognitive decline. *Journal of Experimental Psychology: General*, *145*, 284–297.

Bonawitz, E. B., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*, 215–234.

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., & Woolard, J. (2010). Age differences in affective decision making as indexed by performance on the Iowa gambling task. *Developmental Psychology*, *46*, 193–207.

Davidson, D. (1991). Developmental differences in children's search of predecisional information. *Journal of Experimental Child Psychology*, *52*, 239–255.

Davidson, D. (1996). The effects of decision characteristics on children's selective search of predecisional information. *Acta Psychologica*, *92*, 263–281.

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychological Science*, *27*, 848–858.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*, 1062–1068.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science*, *24*, 87–92.

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, USA*, *114*, 7892–7899.

Hagen, J. W., & Hale, G. A. (1973). The development of attention in children. *ETS Research Report Series*, *1973*, i–37. doi:10.1002/j.2333-8504.1973.tb00453.x

Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision-making. *Current Opinion in Behavioral Sciences*, *5*, 108–115.

Josef, A. K., Richter, D., Samanez-Larkin, G. R., Wagner, G. G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology*, *111*, 430–450.

Keresztes, A., Bender, A., Bodammer, N., Lindenberger, U., Shing, Y. L., & Werkle-Bergner, M. (2017). Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. *Proceedings of the National Academy of Sciences, USA*, *114*, 9212–9217.

Klahr, D. (1982). Nonmonotone assessment of monotone development: An information processing analysis. In S. Strauss & R. Stavy (Eds.), *U-shaped behavioral growth* (pp. 63–86). New York, NY: Academic Press.

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*, 284–299.

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.

Mata, R., Wilke, A., & Czienskowski, U. (2013). Foraging across the life span: Is there a reduction in exploration with aging? *Frontiers in Neuroscience*, *7*, Article 53. doi:10.3389/fnins.2013.00053

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., . . . Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215.

Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S.-J. (2016). The computational development of reinforcement learning during adolescence. *PLOS Computational Biology*, *12*(6), Article e1004953. doi:10.1371/journal.pcbi.1004953

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, *2*, 176–186.

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Riedmiller, M., Hafner, R., Lampe, T., Neunert, M., Degrave, J., Van de Wiele, T., . . . Springenberg, J. T. (2018). *Learning by playing: Solving sparse reward tasks from scratch*. Retrieved from arXiv: https://arxiv.org/abs/*1802*.10567

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2017). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 927–943.

Schulz, L. (2015). Infants explore the unexpected. *Science*, *348*, 42–43.

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

Sollich, P. (1999). Learning curves for Gaussian processes. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems* (Vol. 11, pp. 344–350). Cambridge, MA: MIT Press.

Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology: General*, *146*, 155–164.

Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. (2009). *Gaussian process optimization in the bandit setting: No regret and experimental design*. Retrieved from arXiv: https://arxiv.org/abs/0912.3995

Turing, A. (1950). Computing intelligence and machinery. *Mind*, *59*, 433–460.

Tymula, A., Belmaker, L. A. R., Roy, A. K., Ruderman, L., Manson, K., Glimcher, P. W., & Levy, I. (2012). Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences, USA*, *109*, 17135–17140.

White, J. M. (2013). *The role of delayed consequences in human decision-making* (Unpublished doctoral dissertation). Princeton University, Princeton, NJ.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*, 2074–2081.

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.