Supplementary Information

1 Experimental Stimulus Selection

The experimental decision task presents a multi-armed bandit task with 225 book options arranged in a 15x15 grid. We selected these 225 books in the following way: (1) we created a subsample of books with an average rating higher than 4 and more than 20 ratings in the GoodReads metadata, (2) we further subsampled books to include books that have a synopsis longer than 600 characters and shorter than 1200 characters, (3) we randomly sampled 25 books for each genre in a genre list consisting 18 common book genres, (4) the final selections were made based on synopsis quality.

2 Experimental Stimulus Validation

We collected perceived pairwise similarities among 22 sampled books from 248 participants recruited from Prolific. We randomly drew 15 pairs of book synopses for each participant and asked participants to read the synopses and evaluate the similarities between the book pairs. For each pair of these 22 books, we averaged participants' perceived similarity responses (ranging from 1 to 9) and constructed a 22x22 distance matrix. In addition, we also constructed another two 22x22 distance matrices with distance metrics calculated as the Euclidean distance between semantic embedding vectors of the 22 books and the distance metrics calculated as the Euclidean distance between 22 books in the grid space. We compared the two distance matrices and evaluated the association between these two distance metrics using the Mantel test [1]. We conducted the Mantel test using 10,000 permutations on Pearson correlation metrics and conducted two-tailed significance tests. The results are shown in Supplemental Table 1 and Supplemental Figure 1.

Table 1 Mantel test results to evaluate the association among different distance metrics.

Pair of distance metrics	r	Z	Empirical p-value
Perceived distance - Semantic Embedding distance	0.546	7.247	<0.001
Perceived distance - Grid distance	0.410	6.199	<0.001
Semantic Embedding - Grid distance	0.543	8.188	<0.001

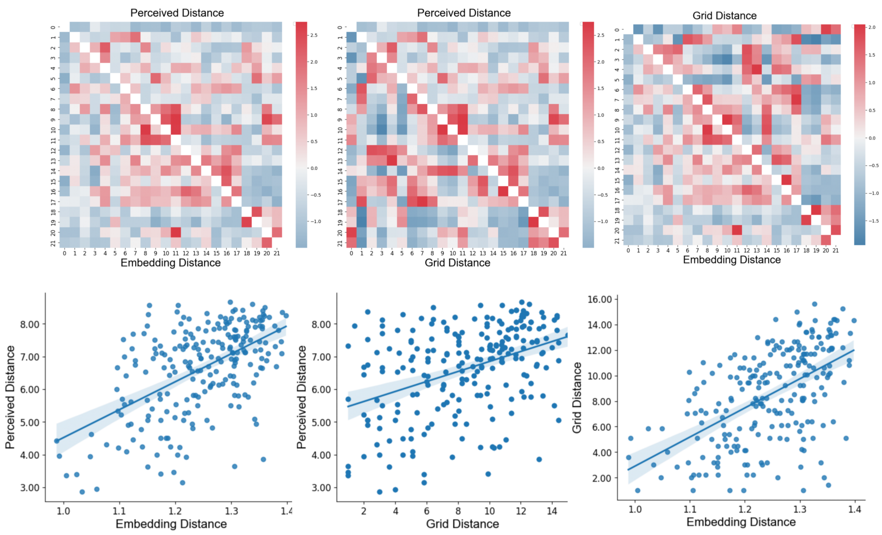

Fig. 1 Upper Panel: Using different distance metrics, the heatplots of the pairwise distance between sampled 22 books. The upper matrix and lower matrix encode different distance metrics. Color encodes normalized distance metrics. Bottom Panel: The regression plots of pairwise associations among three distance metrics. Columns represent different pairs of distance metrics: Semantic Embedding Distance - Subject Perceived Distance (Left), Grid Space Distance - Subject Perceived Distance (Middle), and Semantic Embedding Distance - Grid Space Distance (Right).

3 Clustering tendency of books in semantic embedding space

To measure to what extent books are clustered in the semantic embedding space, we calculated the Hopkins' Statistic as a measure of the clustering tendency of all books [2]. In principle, Hopkins' Statistic compares the distance between a book and its nearest neighbor to the distance between this book and a randomly selected book. The calculated Hopkins' Statistic H equals 0.755, indicating a clustering tendency at the 90% confidence level.

4 Learning to Choose Better Books or Enjoy Books More Over Time?

We found that people choose more favorable books over time, as shown by the increasing trend of book ratings as the number of books read increases. However, alternative explanations exist such that people are not learning to choose better books but are just becoming more favorable toward books as a function of time. This phenomenon, known as coherency maximization [3], is common in purchase behavior such that the more people purchase a product, the more they will like it. To test this alternative explanation, we combined two independent datasets (people's selection histories and GoodReads rating data) and found that there is a positive association between the average rating of books (scrapped from GoodReads) and people's number of past reads ($r = 0.034, p < 0.001, 99\%CI = [0.032, 0.036]$), shown in Supplemental Figure 2. Thus, we show that people are not just increasingly liking the selected book but are learning to select better books.

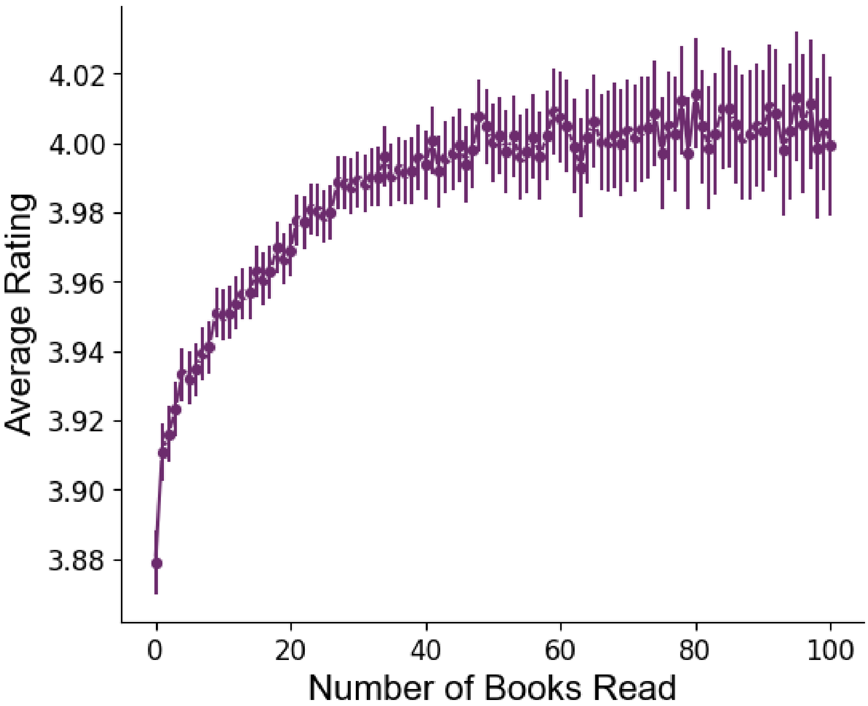

Fig. 2 The lineplot of book average rating (scrapped from GoodReads metadata) by the number of books read in the past.

5 Sensitivity Analysis: Mixed Effect Model for Exploration Distance and Reading Enjoyment

We conducted a sensitivity analysis using a mixed effect model for the exploration distance in the real-world dataset, with a random intercept for each subject in the dataset. Our main analysis found that people's explore distance is associated with (1) the number of past reads, (2) the preceding reading enjoyment, (3) the number of following book reviews, and (4) the variance of following book reviews. Thus in the mixed effect regression model, we specified fixed effect variables as these four variables. Moreover, considering a potential autocorrelation effect in the sequence of explore distances, we added the preceding explore distance as an additional fixed effect variable in our mixed regression model (Equation S1). Here, we report the model fitting regression coefficient estimates in Supplemental Table 2.

$$ExploreDistance_{t,j} \sim \beta_{0,j} + \beta_1 t + \beta_2 ExploreDistance_{t-1,j} + \beta_3 Enjoyment_{t-1,j} + \beta_4 VarianceRating_{t,j} + \beta_5 log(RatingSize_{t,j}) + \epsilon_{t,j} \quad (1)$$

Similarly, we fit the mixed effect model for exploration distance in the experimental dataset. Note that the number and the variance of book reviews were not shown to the participants, we dropped these two variables in the model for experimental data, while adding people's curiosity measures to the model. Thus, we specified the mixed effect model with a random intercept for each subject in the dataset and fixed effect variables as the sequential order of the exploration, the preceding exploration distance, preceding reading enjoyment, and the five dimensions of trait curiosity as independent variables (Equation S2). We report the model fitting regression coefficient estimates in Supplemental Table 3.

$$ExploreDistance_{t,j} \sim \beta_{0,j} + \beta_1 t + \beta_2 ExploreDistance_{t-1,j} + \beta_3 Enjoyment_{t-1,j} + \beta_4 CuriosityJE_j + \beta_5 CuriosityTS_j + \beta_6 CuriosityDS_j + \beta_7 CuriosityST_j + \beta_8 CuriositySC_j + \epsilon_{t,j} \quad (2)$$

Table 2 Result of random intercept mixed effect model regression on explore distance in the real-world dataset.

	Estimate	SE	Z	$P > (\|Z\|)$
Fixed Intercept	0.981	0.001	771.288	< 0.001
Sequential order	0.000	0.000	9.000	< 0.001
Preceding explore distance	0.163	0.001	216.567	< 0.001
Preceding reading enjoyment	-0.005	0.000	-37.621	< 0.001
Current rating variance	0.035	0.000	78.740	< 0.001
Current log number of ratings	-0.002	0.000	-41.940	< 0.001

Furthermore, our main analysis found people's explore distance is associated with their following reading enjoyment, and there exists an interaction effect between joyous exploration and explore distance on people's following reading enjoyment. Thus, in this sensitivity analysis, we fitted a mixed-effect regression model on people's reading enjoyment with independent variables including (1) preceding explore distance, (2) joyous exploration, and (3) an interaction effect between joyous exploration and preceding explore distance. Similar to the mixed effect model for the explore distance mentioned above, we specified an additional fixed effect variable as the sequential order of the exploration and the preceding reading enjoyment to account for the potential trend effect and autocorrelation effect (Equation S3). We report the results in Supplemental Tables 4.

Table 3 Result of random intercept mixed effect model regression on explore distance in the experimental dataset.

	Estimate	SE	Z	$P > (\|Z\|)$
Fixed Intercept	8.156	0.740	11.024	< 0.001
Sequential order	-0.118	0.020	-6.015	< 0.001
Preceding explore distance	0.278	0.016	17.328	< 0.001
Preceding reading enjoyment	-0.577	0.034	-16.925	< 0.001
Joyous Exploration	0.287	0.121	2.365	0.019
Thrill Seeking	0.158	0.108	1.469	0.143
Deprivation sensitivity	-0.102	0.103	-0.992	0.322
Stress tolerance	-0.141	0.092	-1.533	0.127
Social Curiosity	-0.120	0.084	-1.422	0.157

$$\begin{aligned} Enjoyment_{t,j} \sim \beta_{0,j} + \beta_1 t + \beta_2 Enjoyment_{t-1,j} + \beta_3 ExploreDistance_{t-1,j} \\ + \beta_4 CuriosityJE_j + \beta_5 ExploreDistance_{t-1,j} * CuriosityJE_j \\ + \epsilon_{t,j} \end{aligned} \tag{3}$$

Table 4 Result of random intercept mixed effect model regression on reading enjoyment in the experimental dataset.

	Estimate	SE	Z	$P > (\|Z\|)$
Fixed Intercept	3.721	0.437	8.512	< 0.001
Sequential order	0.042	0.009	4.824	< 0.001
Preceding enjoyment	0.075	0.019	3.935	< 0.001
Preceding explore distance	-0.100	0.037	-2.687	0.007
Joyous Exploration	0.183	0.084	2.191	0.028
Preceding explore distance * Joyous Exploration	0.016	0.007	2.142	0.032

6 Computational Modeling

Following previous studies[4, 5], we constructed two computational models: a Gaussian Process (GP) regression model for the reward generalization mechanism and an Upper Confidence Bound (UCB) model for the directed exploration mechanism.

6.1 Gaussian process regression

In principle, GP[6] formalizes a value function f that takes the options' semantic features x in a multi-dimensional space (semantic embeddings in real-world data and grid embedding in experiment data) as inputs and outputs a scalar value as the expected reward. This function is modeled as a multivariate Gaussian:

$$f \sim GP(m(x), k(x, x')) \tag{4}$$

$$m(x) = E[f(x)] \tag{5}$$
$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \tag{6}$$

where $m(x)$ denotes the expected value for the value function of option x, and $k(x, x')$ encodes the value covariance between two options x and x'.

Consider a sequence of observations of book selections $\boldsymbol{X} = [x_1, x_2, ..., x_t]$ and their corresponding rewards $\boldsymbol{y} = [y_1, y_2, ..., y_t]$ from time 1 to time t. GP regression uses the Bayesian principle to generate the posterior predictions of a target option x^*, by computing its posterior mean $m(x^*)$ and variance $v(x^*)$ conditional on observations $\boldsymbol{X}, \boldsymbol{y}$:

$$m(x^*|\boldsymbol{X}, \boldsymbol{y}) = \boldsymbol{k_*}^T(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{y} \tag{7}$$
$$v(x^*|\boldsymbol{X}, \boldsymbol{y}) = \boldsymbol{k_*}^T(\boldsymbol{K} + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{k_*}^T \tag{8}$$

where $\boldsymbol{k_*}$ denotes the $1 * t$ covariance matrix between input observations $\boldsymbol{X}$ and the target option x^*, $\boldsymbol{K}$ denotes the $t * t$ covariance matrix of the input observations $\boldsymbol{X}$, and σ^2 denotes the assumed noise from the observations. Here, we specify $\sigma^2 = 1$.

The covariance between two options is defined by a Radical Kernel function so that the posterior estimation depends on the Euclidean distance between target options and the input observations:

$$k(x, x') = exp(\frac{||x - x'||^2}{2\lambda^2}) \tag{9}$$

In this way, the GP regression and the kernel function together formalize the reward generalization mechanism by assuming that similar options will generate similar reward outcomes. The length-scale parameter λ in the kernel function regulates this generalization process by controlling the smoothness of the generalization function, in a way such that $\lambda \to 0$ leads to zero generalization and independent value estimation among options, whereas $\lambda \to \infty$ leads to maximum generalization, such that the dependency of value estimation is linear to feature distances.

6.2 Value Function and Decision Function

The GP regression model simply considers that the value for a target option $Q_{GP}(x^*)$ equals its expected reward $m(x^*)$ as:

$$Q_{GP}(x^*) = m(x^*) \tag{10}$$

Conversely, the UCB model specifies the value for a target option $Q_{UCB}(x^*)$ as a weighted sum of the expected reward $m(x^*)$ and the square root of the variance of the reward estimation $v(x^*)$:

$$Q_{UCB}(x^*) = m(x^*) + \beta\sqrt{v(x^*)} \tag{11}$$

where the exploration bonus parameter β controls the extent of directed exploration, with higher β leading to a stronger bias towards options with high uncertainty.

Next, we specify the decision function for both models as a softmax function, which maps the estimated value for N available options to the probabilities of choosing an option x_i as $p(x_i)$:

$$p(x_i) = \frac{exp(Q(x_i)/\tau)}{\sum_j^N exp(Q(x_j)/\tau)} \tag{12}$$

where the random temperature parameter τ in this softmax controls the randomness in the probabilistic mechanism. $\tau \to 0$ leads to zero randomness, such that the highest-valued option is always chosen, whereas $\tau \to \infty$ leads to maximum randomness with a uniform probability of selecting any option.

7 A Challenge for Comparing Computational Models with Real-world Data

We recognize model performance comparisons depend on the relative exploration or exploitation tendencies between observed choices and alternatives. Consider, an empirical choice was observed and used to estimate the performance of models with or without directed exploration. If all the alternative options were high in estimated uncertainty, the observed empirical choice would become relatively more exploitative, thus favoring a model without directed exploration. Similarly, an exploitive option space for the available alternatives will bias the model comparison, favoring the model with a directed exploration mechanism.

As a result, information on the option space is vital for an unbiased comparison between exploration and exploitation models. However, we do not know such information for real-world data. Consistent with prior research [7], we specified an artificial alternative option with averaged features of all available options, but this fixed, arbitrary, and small option space inevitably introduces opaque biases in our comparison results for real-world data. We note that our results from experiment data are not subject to this issue because the alternative option space is always known. Thus, our parallel analysis of real-world and experimental data helps us overcome this challenge.

8 Multiple Regression on Exploration Parameters

We applied multiple regression models on the exploration parameters (i.e., τ, β) with curiosity dimensions as independent variables while controlling for participants' age, as specified in Equation S13. The result of the regression model is reported in Supplementary Table 5.

$$\begin{aligned} \tau, \beta \sim\ & \beta_0 + \beta_1 Age + \beta_2 CuriosityJE + \beta_3 CuriosityTS + \\ & \beta_4 CuriosityDS + \beta_5 CuriosityST + \beta_6 CuriositySCj + \epsilon \end{aligned} \tag{13}$$

In addition, we plotted the bi-variate pairwise correlation matrix among the dependent variables and independent variables in our specified regression model in Supplemental Figure 3.

Table 5 Result of multiple regression on decision parameters.

	Bonus β				Temperature τ			
	Estimate	SE	$t(238)$	$P > (\lvert t \rvert)$	Estimate	SE	$t(238)$	$P > (\lvert t \rvert)$
Intercept	-1.405	1.340	-1.049	0.295	-0.9227	0.634	-1.456	0.147
Age	-0.007	0.016	-0.442	0.659	0.0027	0.007	0.359	0.720
Joyous-Exploration	**0.605**	**0.214**	**2.824**	**0.005**	**-0.2011**	**0.101**	**-1.983**	**0.049**
Thrill-Seeking	-0.150	0.195	-0.771	0.441	**0.2247**	**0.092**	**2.440**	**0.015**
Deprivation-Sensitivity	-0.150	0.183	-0.816	0.416	-0.1213	0.087	-1.399	0.163
Stress-Tolerance	0.039	0.169	0.228	0.820	-0.0237	0.080	-0.296	0.768
Social-Curiosity	0.021	0.149	0.137	0.891	0.0647	0.071	0.915	0.361

Note: Statistics of significant hypothesis testing results were bolded in the table.

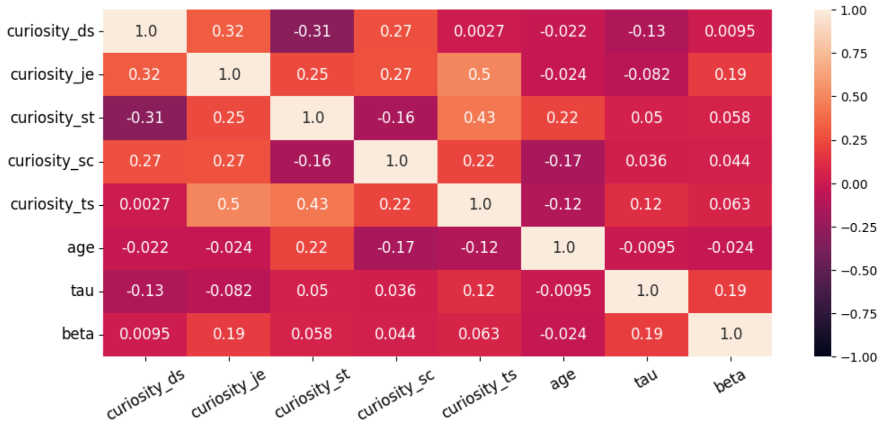

Fig. 3 The bivariate Pearson correlation matrix of individual-level variables.

9 Sensitivity Analysis on Multiple Regression

We conducted a sensitivity analysis for the multiple regression on the exploration parameters after the removal of the non-significant variables. The results are reported in Supplemental Table 6 for exploration bonus β and Supplemental Table 7 for random temperature τ.

10 Sensitivity Analysis with Censored Regression

We recognized that a considerable portion of the parameter estimation for the exploration bonus β reaches the boundary close to 0. This issue might hamper the regression coefficient estimation due to a non-normal error distribution. Thus, we fit a censored regression on exploration bonus β, which assumes the dependent variable is censored

Table 6 Result of regression model on log exploration bonus β after removal of non-significant variables.

	Estimate	SE	$t(243)$	$P > (\lvert t \rvert)$
Intercept	-2.039	0.841	-2.423	0.016
Joyous Exploration	0.491	0.167	2.945	0.004

Table 7 Result of multiple regression on log random temperature τ after removal of non-significant variables.

	Estimate	SE	$t(242)$	$P > (\lvert t \rvert)$
Intercept	-1.0588	0.402	-2.631	0.009
Joyous Exploration	-0.2385	0.091	-2.613	0.010
Thrill Seeking	0.2418	0.081	2.974	0.003

at boundaries near 0. The result remains consistent with the multiple regression results and reported in Supplemental Table 8.

Table 8 Result of censored regression on log exploration bonus β.

	Estimate	SE	$t(237)$	$P > (\lvert t \rvert)$
Intercept	-1.957	1.473	-1.329	0.184
Age	-0.006	0.017	-0.354	0.724
Joyous Exploration	**0.692**	**0.235**	**2.944**	**0.003**
Thrill Seeking	-0.158	0.213	-0.743	0.458
Deprivation Sensitivity	-0.169	0.201	-0.839	0.402
Stress Tolerance	0.027	0.186	0.144	0.885
Social Curiosity	0.045	0.163	0.279	0.781

Note: Statistics of significant hypothesis testing results were bolded in the table.

11 Parameter Estimation Stability

We estimated the parameters using a global optimization method—the differential evolution algorithm. This global optimization method is non-deterministic, so we repeated the parameter estimation 100 times and took the average as our final parameter estimate. Supplemental Figure 4 plots the within-subject variance of these 100 repetitions of parameter estimations and shows that there is very small within-subject variability for both parameter estimates, thus indicating that our parameter estimation method is stable.

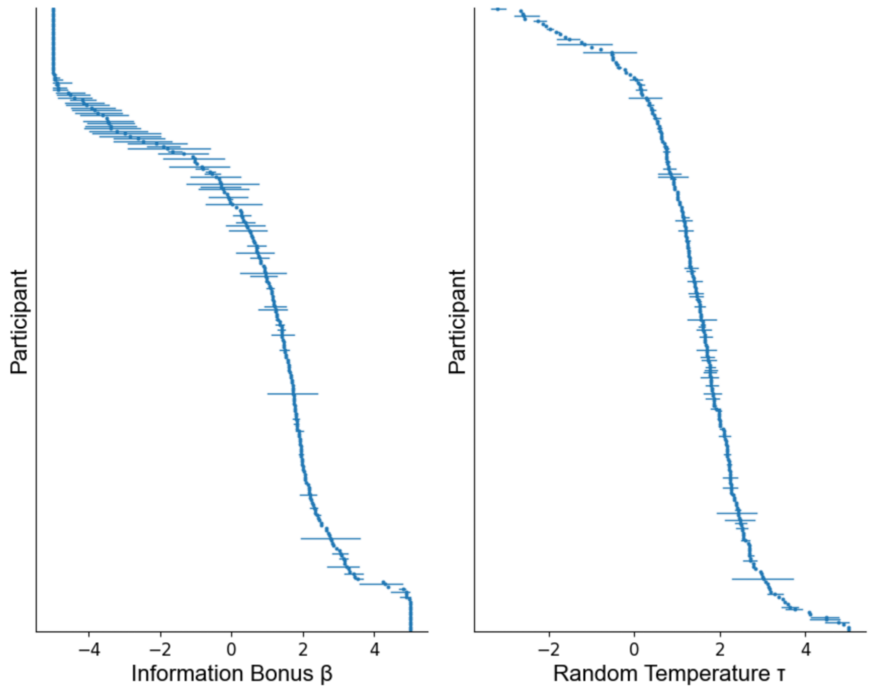

Fig. 4 Parameter estimations across different optimization repetition and different participants. The y-axis encodes the participants ordered by a mean estimate of the parameter in descending order, while the x-axis encodes the exploration parameter for exploration bonus β (left) and random temperature τ (right). The horizontal line encodes the 99% confidence interval of 100 repetitions of parameter estimations.

12 Cronbach's α for Curiosity Dimensions

To estimate the validity of the five-dimension curiosity scale, we measured Cronbach's α statistic for each of the five dimensions including deprivation sensitivity, joyous exploration, stress tolerance, social curiosity, and thrill-seeking. The results are shown in Table 9 below. All dimensions of curiosity have a Cronbach's α larger than 0.75, indicating good reliability of the scale.

Table 9 Cronbach's α for the five curiosity dimensions

	Cronbach's α	95% Confidence Interval
Deprivation Sensitivity	0.822	[0.788, 0.853]
Joyous Exploration	0.788	[0.742, 0.827]
Stress Tolerance	0.854	[0.823, 0.881]
Social Curiosity	0.833	[0.797, 0.864]
Thrill Seeking	0.791	[0.747, 0.830]

References

[1] Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220 (1967).

[2] Lawson, R. G. & Jurs, P. C. New index for clustering tendency and its application to chemical problems. *Journal of Chemical Information and Computer Sciences* **30**, 36–41 (1990).

[3] Riefer, P. S., Prior, R., Blair, N., Pavey, G. & Love, B. C. Coherency-maximizing exploration in the supermarket. *Nature Human Behaviour* **1**, 1–4 (2017). URL https://www.nature.com/articles/s41562-016-0017.

[4] Schulz, E. *et al.* Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences* **116**, 13903–13908 (2019).

[5] Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D. & Meder, B. Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour* **2**, 915–924 (2018).

[6] Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (The MIT Press, 2005). URL https://direct.mit.edu/books/book/2320/Gaussian-Processes-for-Machine-Learning.

[7] Schulz, E. *et al.* Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences* **116**, 13903–13908 (2019). URL https://www.pnas.org/content/116/28/13903.