

Supplementary Materials for “Curse of Simplicity in High-Dimensional Environments”

Yiming Wang^{1*}, Eeshan Hasan¹, Vladimir M. Sloutsky¹,
Brandon M. Turner¹

^{1*}Department of Psychology, The Ohio State University, 1835 Neil
Avenue, Columbus, 43210, Ohio, United States.

*Corresponding author(s). E-mail(s): wang.18319@osu.edu;
Contributing authors: eeshan.hasan@gmail.com; sloutsky.1@osu.edu;
turner.826@gmail.com;

1 Model Implementation

1.1 Model parameterization

We assume that varying the value of p within the range $0 < p < 1$ allows the model to capture different profiles of attentional constraint. Figure 1 shows simulated values of the p -norm of sampling probabilities in a two-dimensional attention space. Let the vector of sampling probabilities be

$$\mathbf{p}_{\text{samp}} = [p_{\text{samp}_1}, p_{\text{samp}_2}]. \quad (1)$$

The sampling cost is defined as the ℓ_p quasi-norm of this vector:

$$\|\mathbf{p}_{\text{samp}}\|_p = (p_{\text{samp}_1}^p + p_{\text{samp}_2}^p)^{1/p}. \quad (2)$$

When $0 < p < 1$, the equal-value contours of $\|\mathbf{p}_{\text{samp}}\|_p$ become concave. As p approaches zero, for example when $p = .1$, the cost structure favors either uniformly low sampling probabilities across dimensions or concentrated sampling on a single dimension. This property captures a dimension-reduction tendency: broadly distributed sampling becomes increasingly costly, whereas selective sampling along one dimension is relatively favored. In this sense, small values of p can produce an effect similar to attentional competition, but without imposing a strict sum-to-constant assumption in which increasing attention to one dimension necessarily requires an equivalent decrease in the amount of total attention to other dimensions. Thus, allowing p to vary within $(0, 1)$ provides a flexible way to model different degrees of the simplicity goal during learning. Consistent with this motivation, [Galdo, Weichart, Sloutsky, and Turner \(2022\)](#) showed that, among models with $p > 0$, models imposing a structural constraint with $p < 1$ provided the best account of attentional allocation. We therefore treated p as a free parameter bounded between 0 and 1 during model fitting.

1.2 Model initialization

Before the first training trial, the model memory was initialized with four pseudo-exemplars. Two pseudo-exemplars contained feature value 0 on every dimension and two contained feature value 1 on every dimension. Each stimulus configuration was represented once in each response category, such that the initial category labels were [A, B, B, A] for the four pseudo-exemplars. Consequently, the pseudo-exemplars contributed symmetric evidence to the two categories and did not initially favor either response category. All four pseudo-exemplars were assigned the common initial latent attention value α_0 across dimensions. Following the onset of training, observed stimuli were added sequentially to the model memory consistent with the model-learning procedure described in the Main Text.

2 Derivation of loss gradient for error–effort model

For participant or trial i , the loss function was defined as the weighted sum of the cross-entropy loss—error term—and the sampling-cost—effort term:

$$\mathcal{L}^{(i)} = \text{CE}^{(i)} + w \left\| \mathbf{psamp}^{(i)} \right\|_p \quad (3)$$

$$= -\log P^{(i)}\left(F^{(i)}\right) + w \left\| \mathbf{psamp}^{(i)} \right\|_p, \quad (4)$$

where $P^{(i)}(F^{(i)})$ is the model-predicted probability assigned to the correct category, i.e., probability of responding correctly, $\mathbf{psamp}^{(i)}$ is the vector of sampling probabilities, w controls the weight of the sampling-cost term, and p determines the strength of the ℓ_p -based attentional constraint. The derivative of the loss with respect to the latent attention parameter $\alpha_j^{(i)}$ is

$$\frac{\partial \mathcal{L}^{(i)}}{\partial \alpha_j^{(i)}} = -\frac{\partial}{\partial \alpha_j^{(i)}} \log P^{(i)}\left(F^{(i)}\right) + w \frac{\partial}{\partial \alpha_j^{(i)}} \left\| \mathbf{psamp}^{(i)} \right\|_p. \quad (5)$$

2.1 Derivative of error term

Let the category evidence for category c be

$$A_c^{(i)} = \sum_n a\left(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}\right) \mathbb{I}\left(F^{(n)} = c\right), \quad (6)$$

where $a(\mathbf{e}^{(i)}, \mathbf{x}^{(n)})$ denotes the similarity between the current stimulus representation $\mathbf{e}^{(i)}$ and exemplar $\mathbf{x}^{(n)}$, and $\mathbb{I}(\cdot)$ is an indicator function. The total evidence across categories is

$$A^{(i)} = \sum_k A_k^{(i)} = \sum_k \sum_n a\left(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}\right) \mathbb{I}\left(F^{(n)} = k\right). \quad (7)$$

Thus, the predicted category probability is

$$P^{(i)}(c) = \frac{A_c^{(i)}}{A^{(i)}}. \quad (8)$$

For the correct category $F^{(i)}$, the derivative of the log predicted probability is

$$\frac{\partial}{\partial \alpha_j^{(i)}} \log P^{(i)}\left(F^{(i)}\right) = \frac{1}{A_{F^{(i)}}^{(i)}} \frac{\partial A_{F^{(i)}}^{(i)}}{\partial \alpha_j^{(i)}} - \frac{1}{A^{(i)}} \frac{\partial A^{(i)}}{\partial \alpha_j^{(i)}}. \quad (9)$$

The two derivatives in the expression above are

$$\frac{\partial A_{F^{(i)}}^{(i)}}{\partial \alpha_j^{(i)}} = \sum_n \frac{\partial}{\partial \alpha_j^{(i)}} a\left(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}\right) \mathbb{I}\left(F^{(n)} = F^{(i)}\right), \quad (10)$$

and

$$\frac{\partial A^{(i)}}{\partial \alpha_j^{(i)}} = \sum_k \sum_n \frac{\partial}{\partial \alpha_j^{(i)}} a(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) \mathbb{I}(F^{(n)} = k). \quad (11)$$

The similarity function was defined as the product of dimension-specific similarities,

$$a(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) = \prod_m a_m(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}), \quad (12)$$

where

$$a_m(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) = \exp \left[-\delta f(\alpha_m^{(i)}) d_m(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) g(\alpha_m^{(n)}) \right]. \quad (13)$$

Only the j th dimension depends on $\alpha_j^{(i)}$. Therefore,

$$\frac{\partial}{\partial \alpha_j^{(i)}} a(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) = a(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) \left[-\delta d_j(\mathbf{e}^{(i)}, \mathbf{x}^{(n)}) g(\alpha_j^{(n)}) \frac{\partial}{\partial \alpha_j^{(i)}} f(\alpha_j^{(i)}) \right]. \quad (14)$$

With the logistic transformation,

$$f(\alpha_j^{(i)}) = \frac{1}{1 + \exp(-\alpha_j^{(i)})}, \quad g(\alpha_j^{(n)}) = \frac{1}{1 + \exp[-(b_{0,j} + b_1 \alpha_j^{(n)})]}. \quad (15)$$

Thus, the derivative of $f(\alpha)$ is

$$\frac{\partial}{\partial \alpha_j^{(i)}} f(\alpha_j^{(i)}) = f(\alpha_j^{(i)}) [1 - f(\alpha_j^{(i)})]. \quad (16)$$

2.2 Derivative of effort term

The effort term was defined through sampling-cost as

$$\|\mathbf{psamp}^{(i)}\|_p = \left(\sum_m [\mathbf{psamp}_m^{(i)}]^p \right)^{1/p}. \quad (17)$$

The derivative with respect to $\alpha_j^{(i)}$ is

$$\frac{\partial}{\partial \alpha_j^{(i)}} \|\mathbf{psamp}^{(i)}\|_p = \frac{\partial}{\partial \alpha_j^{(i)}} \left(\sum_m [\mathbf{psamp}_m^{(i)}]^p \right)^{1/p} \quad (18)$$

$$= \left(\sum_m [\mathbf{psamp}_m^{(i)}]^p \right)^{1/p-1} [\mathbf{psamp}_j^{(i)}]^{p-1} \frac{\partial \mathbf{psamp}_j^{(i)}}{\partial \alpha_j^{(i)}}. \quad (19)$$

The sampling probability for dimension j was linked to latent attention by

$$\text{psamp}_j^{(i)} = \frac{1}{1 + \exp\left[-\left(b_{0,j} + b_1 \alpha_j^{(i)}\right)\right]}. \quad (20)$$

Therefore,

$$\frac{\partial \text{psamp}_j^{(i)}}{\partial \alpha_j^{(i)}} = b_1 \cdot \text{psamp}_j^{(i)} \left(1 - \text{psamp}_j^{(i)}\right). \quad (21)$$

Substituting this expression into the derivative of the sampling-cost term gives

$$\frac{\partial}{\partial \alpha_j^{(i)}} \left\| \mathbf{psamp}^{(i)} \right\|_p = b_1 \left(\sum_m \left[\text{psamp}_m^{(i)} \right]^p \right)^{1/p-1} \left[\text{psamp}_j^{(i)} \right]^p \left(1 - \text{psamp}_j^{(i)}\right). \quad (22)$$

3 Supplementary Tables

Table 1: Detailed learner-type distribution across dimensionality conditions.

Dimensionality condition	Total N	D learners	P learners	I learners	Mixed learners
Three-dimensional condition	23	17 (73.9%)	1 (4.3%)	0 (0.0%)	5 (21.7%)
Five-dimensional condition	19	7 (36.8%)	10 (52.6%)	0 (0.0%)	2 (10.5%)
Seven-dimensional condition	19	7 (36.8%)	9 (47.4%)	1 (5.3%)	2 (10.5%)

Note. Learners were classified based on each participant's average sampling proportions during the test phase. Participants were classified as D learners if at least 50% of their sampling time was allocated to the deterministic feature. In the three-dimensional condition, participants were classified as P or I learners if at least 50% of their sampling time was allocated to the probabilistic or irrelevant feature(s), respectively. In the five- and seven-dimensional conditions, participants were classified as P learners if the summed sampling proportion across all probabilistic features was at least 50% and the largest single-feature sampling proportion occurred on a probabilistic feature. Participants who did not meet any of these criteria were classified as Mixed learners, who most likely used multiple features for decisions.

Table 2: Parameter estimates of the error-effort model in the three-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,P}$	p
1	8.470	2.083	23.479	-28.034	9.588	0.641	1.000
2	4.492	62.733	0.000	35.322	17.804	17.894	0.026
3	1.640	3.399	17.833	83.246	-25.068	-18.179	0.999
4	1.342	3.439	4.094	-2.473	-1.198	0.808	0.381
5	1.206	2.377	21.782	4.508	-6.381	-3.373	0.094
6	1.143	2.876	3.415	-1.668	-0.490	0.065	0.546
7	1.106	1.742	1.541	1.589	2.376	2.777	0.254
8	1.024	2.211	6.514	-3.247	-1.036	-0.343	0.427
9	1.018	13.289	7.962	9.502	0.197	-4.877	1.000
10	0.949	4.032	24.261	-0.928	2.685	-0.786	0.172
11	0.947	3.497	9.660	-6.259	-5.245	1.024	0.709
12	0.928	3.043	4.462	-2.947	0.381	0.250	0.913
13	0.897	3.348	22.962	-17.004	-14.830	31.806	1.000
14	0.726	1.121	7.439	-1.565	0.234	0.394	0.413
15	0.670	3.933	51.878	-5.165	-1.924	8.611	0.053
16	0.660	3.913	4.728	-1.245	0.644	-0.156	0.483
17	0.628	3.007	9.143	-2.090	0.030	0.872	0.420
18	0.403	5.534	67.910	5.370	2.911	-18.446	0.423
19	0.114	1.311	6956.364	-5.615	4.658	2.068	0.035
20	-6.556	0.000	1.528	10.421	10.335	9.141	1.000
21	-5.590	1.282	0.239	1.101	0.362	0.932	0.868
22	-3.656	100.029	1.986	11.199	12.687	12.045	0.207
23	-2.207	38.206	3.350	3.403	19.637	8.503	0.968

Note. $w = 0.05$ during the model fitting.

Table 3: Parameter estimates of the error–effort model in the five-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,Pa}$	$b_{0,Pb}$	$b_{0,Pc}$	p
1	5.705	0.032	0.523	-0.902	-0.943	0.605	0.547	-0.150	0.327
2	5.654	0.011	0.655	-1.850	-1.610	-2.521	-1.440	4.235	0.480
3	4.222	1.379	1.606	2.057	-1.028	-1.014	-0.927	-1.023	0.250
4	24.696	0.566	0.302	-5.177	-6.605	-4.964	4.155	-6.323	0.299
5	2.910	0.247	0.004	1.196	1.804	1.942	1.348	0.974	0.104
6	2.086	0.335	1.117	-0.642	-0.189	0.239	0.223	-0.332	0.671
7	1.802	1.236	2.572	-1.449	-2.771	-1.785	3.705	-2.467	0.949
8	1.123	0.837	0.332	-0.377	0.891	-0.176	-0.554	-0.157	0.653
9	1.037	0.329	4.174	-0.811	0.320	0.211	0.134	0.008	0.593
10	0.771	3.079	6.107	-0.685	-1.431	-0.078	-0.278	0.535	0.558
11	0.636	1.490	6.341	-1.865	-0.112	-0.034	0.285	-0.717	0.833
12	0.546	4.318	5.948	1.045	-3.547	-1.143	0.159	-0.780	0.971
13	0.505	2.069	6.588	-1.326	-0.925	-0.488	-0.002	-0.618	0.852
14	0.446	2.278	52.031	3.543	-2.274	-2.265	2.919	-3.113	0.170
15	0.278	22.409	0.073	0.392	2.642	1.575	0.450	3.172	0.186
16	0.258	3.807	23.977	-2.281	-1.442	-1.269	-0.453	2.732	1.000
17	-5.798	0.054	0.524	3.155	3.198	2.984	2.789	2.256	0.382
18	-3.253	0.905	0.274	1.899	1.316	1.380	1.709	1.535	0.291
19	-1.314	4.953	0.518	-0.665	0.416	0.197	0.221	0.686	0.440

Note. $w = 0.05$ during the model fitting.

Table 4: Parameter estimates of the error–effort model in the seven-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,Pa}$	$b_{0,Pb}$	$b_{0,Pc}$	$b_{0,Pd}$	$b_{0,Pe}$	p	
1		89.959	0.000	0.328	-27.135	-27.136	-27.136	-27.137	-27.130	112.091	-27.584	1.000
2		8.540	0.013	0.491	-1.370	-0.900	-1.021	-1.330	-1.271	-0.145	-1.484	0.359
3		5.665	0.382	0.541	-2.019	-1.553	0.771	0.095	-0.323	-0.328	-1.395	0.994
4		3.231	17.627	1.399	0.737	-1.114	-0.881	-1.385	-0.581	-0.900	-1.065	0.289
5		3.123	1.194	0.947	-1.621	4.145	-1.566	-1.487	1.464	-1.482	-0.625	0.431
6		2.088	0.667	0.672	-0.114	-1.441	3.796	-0.065	-0.073	-1.570	-1.595	0.457
7		2.053	0.542	0.480	0.156	0.102	0.235	0.090	1.520	2.447	0.925	0.356
8		1.769	0.731	0.991	-0.999	-0.289	0.296	3.760	-0.639	-0.605	-1.522	0.615
9		1.407	0.691	0.939	0.307	1.271	0.390	0.274	1.064	0.261	-0.462	0.401
10		1.387	6.733	2.723	1.252	-1.189	-0.651	-0.355	-0.723	-0.747	-0.857	0.464
11		1.123	23.280	4.426	2.717	-1.119	-0.783	-1.860	-0.766	-0.841	-0.735	0.349
12		1.047	2.786	3.103	0.911	-1.241	-0.372	0.051	-0.838	-0.376	-0.078	0.685
13		0.955	0.547	2.230	-0.980	-0.435	-0.566	-0.820	-0.190	1.209	-0.689	0.688
14		0.865	3.573	2.353	1.637	-0.936	-0.022	-0.410	-0.581	-0.321	-0.774	0.591
15		0.409	2.253	3.000	-0.668	-0.404	0.501	0.136	0.627	1.684	0.813	0.845
16		-37.961	0.000	0.212	7.732	6.971	7.651	7.821	7.083	7.288	7.280	1.000
17		-3.189	2.904	0.372	1.048	0.305	0.469	0.859	0.979	0.168	0.311	0.367
18		-1.298	0.000	6.577	10.651	-2.423	12.030	19.575	12.031	12.032	12.782	0.352
19		-0.981	2.017	0.872	0.004	-1.054	-0.641	-0.016	0.478	0.508	-1.328	0.523

Note. $w = 0.05$ during the model fitting.

Table 5: Parameter estimates of the error–effort model in the three-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,P}$	p	w
1	8.279	2.083	4539.923	-1.008	4.636	0.521	0.718	0
2	4.590	2.960	48.089	-219.090	-217.127	1015.340	1.000	0
3	2.295	3.209	2.208	0.585	-0.702	-0.268	0.401	0.188
4	14.539	2.771	0.005	-0.343	-0.925	-0.416	0.147	1
5	1.841	3.576	6.981	6.342	-7.336	-4.431	0.994	0.989
6	1.613	1.117	2.929	-0.155	0.654	0.860	0.197	0.013
7	1.059	1.593	3.850	-1.623	-0.255	0.407	0.485	0.055
8	1.058	2.378	34.220	8.109	-14.519	-10.036	0.053	0
9	1.046	4.125	3.362	-0.587	1.158	0.153	0.203	0.007
10	1.036	3.527	5.645	-2.033	-0.431	-0.057	0.195	0.005
11	0.914	2.867	3.931	-1.644	-0.322	0.630	0.518	0.033
12	0.909	5.141	27.663	-15.071	8.246	0.253	0.903	0.683
13	0.876	3.786	5.089	-2.558	-0.753	-1.560	0.564	0.058
14	0.729	2.133	4.062	0.470	2.097	2.528	0.495	0.141
15	0.662	12.606	11.646	-2.507	5.056	-4.209	0.985	0.025
16	0.552	3.313	6.120	-0.814	0.148	3.584	0.276	0.007
17	0.450	4.081	42.926	11.591	-2.224	6.014	0.051	0
18	0.233	4.405	4.483	-0.760	3.938	3.958	0.201	0
19	0.228	5.532	119.735	-19.336	25.648	-14.762	0.887	0.169
20	0.109	1.313	100000.000	-9.651	-3.639	-2.070	0.555	0.996
21	0.027	4.904	884.371	-4.284	3.555	-3.444	0.062	0
22	-193.991	0.000	0.291	56.742	56.672	55.553	0.790	1
23	-0.055	0.402	32.576	-1.053	5.231	0.473	0.980	0

Note. w was a free parameter during model fitting.

Table 6: Parameter estimates of the error-effort model in the five-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,Pa}$	$b_{0,Pb}$	$b_{0,Pc}$	p	w
1	7.217	0.288	0.043	1.586	1.791	2.268	2.100	1.741	0.084	0
2	3.490	0.012	1.042	-1.792	-1.558	-2.461	-1.381	9.144	0.466	0.018
3	3.158	0.007	0.417	0.814	0.776	2.178	2.226	1.507	0.156	0
4	20.776	0.000	0.000	1.346	1.865	1.986	1.477	1.069	0.051	1
5	2.043	0.012	2.070	0.288	0.119	0.640	0.199	0.083	0.341	0.020
6	14.192	0.556	1.535	-19.187	-20.662	-18.921	31.138	-20.469	0.999	0.133
7	13.486	0.000	0.443	-4.736	-5.668	-5.543	-5.072	-5.315	1.000	1
8	1.257	1.325	3.198	-1.303	-2.190	-1.214	6.850	-1.888	0.985	0.037
9	1.074	2.002	6.422	-0.675	-0.962	-0.008	-0.721	-0.670	0.468	0.091
10	0.983	1.160	1.217	-0.902	1.483	0.410	-0.471	1.664	0.244	0.002
11	0.877	3.115	5.144	1.035	-1.134	-0.040	-0.129	0.579	0.398	0.021
12	0.560	4.406	14.709	6.289	-3.214	-2.834	-2.632	-0.939	0.816	0.139
13	0.487	3.662	3.106	0.927	0.435	0.490	1.923	0.205	0.170	0
14	0.459	1.015	0.370	-0.271	1.117	0.025	-0.343	0.038	0.717	0.063
15	0.440	2.022	52.810	4.941	-5.104	-5.007	-0.669	-5.857	0.300	0.404
16	0.394	1.021	5.366	-0.392	1.209	0.944	1.048	0.658	0.120	0
17	0.315	1.179	12.065	-1.874	-0.743	-0.962	-0.256	-1.502	0.419	0.003
18	0.265	0.000	0.431	0.235	0.019	-0.257	-0.325	-1.086	1.000	1
19	-2.981	25.497	0.978	1.295	3.073	2.593	2.618	2.992	0.161	0

Note. w was a free parameter during model fitting.

Table 7: Parameter estimates of the error–effort model in the seven-dimensional condition.

ID	α_0	δ	b_1	$b_{0,D}$	$b_{0,I}$	$b_{0,Pa}$	$b_{0,Pb}$	$b_{0,Pc}$	$b_{0,Pd}$	$b_{0,Pe}$	p	w
1	5.347	0.631	0.409	-1.121	-0.726	-0.099	6.584	-1.092	-1.057	-1.868	0.486	0.118
2	5.253	0.379	0.161	0.140	0.614	3.000	2.278	1.884	1.856	0.778	0.946	0
3	3.497	2.474	2.449	36.630	-7.173	-7.089	-7.170	-7.195	-7.009	-7.540	0.393	0.007
4	2.063	1.141	1.505	-1.648	4.907	-1.393	-1.428	1.508	-1.424	-0.755	0.229	0
5	1.644	0.163	1.999	-0.688	0.115	-0.065	-0.376	-0.284	0.760	-0.532	0.921	0.099
6	1.624	2.750	2.839	0.658	-0.207	0.069	-0.443	0.284	0.039	-0.429	0.278	0.026
7	1.473	15.824	2.249	1.265	-0.718	0.029	0.424	-0.210	-0.170	-0.245	0.391	0.037
8	1.354	0.590	1.191	-0.336	-0.042	-0.061	-0.348	0.219	1.718	-0.272	0.512	0.060
9	1.172	1.015	1.388	-0.680	-0.364	-0.161	-0.494	0.864	1.732	0.345	0.373	0.009
10	0.990	30.411	4.932	1.585	-0.972	-0.678	-1.743	-0.637	-0.721	-0.619	0.381	0.066
11	0.697	2.948	5.097	1.009	-2.045	-0.435	0.128	-1.201	-0.437	0.002	0.766	0.026
12	0.543	1.804	2.946	-0.913	-0.288	-0.085	-0.318	0.201	0.944	0.007	0.657	0.025
13	0.507	0.005	0.265	1.875	2.377	1.747	1.733	2.280	1.614	1.892	0.159	0
14	0.114	0.167	11.051	-0.629	1.218	0.635	0.602	0.564	10.645	0.152	1.000	0
15	-26.964	0.037	0.096	2.241	1.481	2.155	2.326	1.587	1.789	1.789	0.715	0
16	-1.285	0.070	0.444	0.562	-0.320	-0.198	0.327	0.487	-0.656	-0.432	0.861	0.720
17	-1.127	0.875	1.280	0.676	-0.609	-0.019	0.469	0.931	1.068	-0.947	0.704	0.048
18	-0.491	0.077	0.000	1.911	-14.375	3.155	7.165	3.143	3.141	3.390	0.067	0.106
19	-0.164	1.403	0.467	1.179	0.358	6.877	1.488	1.481	0.203	0.128	0.314	0.017

Note. w was a free parameter during model fitting.

4 Supplementary Figures

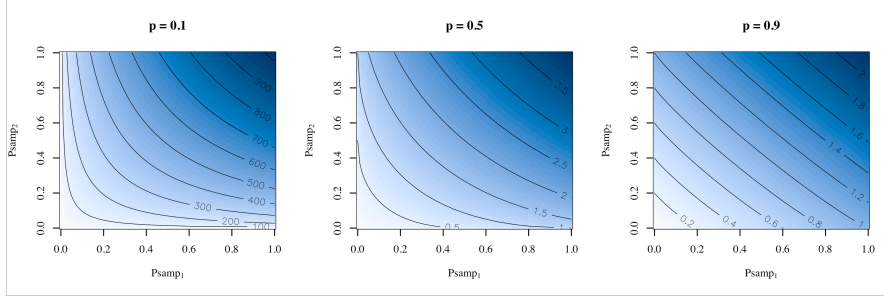


Fig. 1: ℓ_p quasi-norm of sampling probabilities under different p . This figure shows a two-dimensional case, where the vector of sampling probabilities is $\mathbf{psamp} = [\text{Psamp}_1, \text{Psamp}_2]$, and the sampling cost is defined as $\|\mathbf{psamp}\|_p = (\text{Psamp}_1^p + \text{Psamp}_2^p)^{1/p}$. Each panel shows simulated values of $\|\mathbf{psamp}\|_p$ over combinations of Psamp_1 and Psamp_2 under exemplar values of p (left: $p = 0.1$; middle: $p = 0.5$; right: $p = 0.9$). Black lines represent equal-value contours of $\|\mathbf{psamp}\|_p$. Lighter blue shades indicate lower values of $\|\mathbf{psamp}\|_p$. As p decreases from 1 toward 0 (right to left), the contours become increasingly concave, favoring more selective or lower-dimensional sampling profiles.

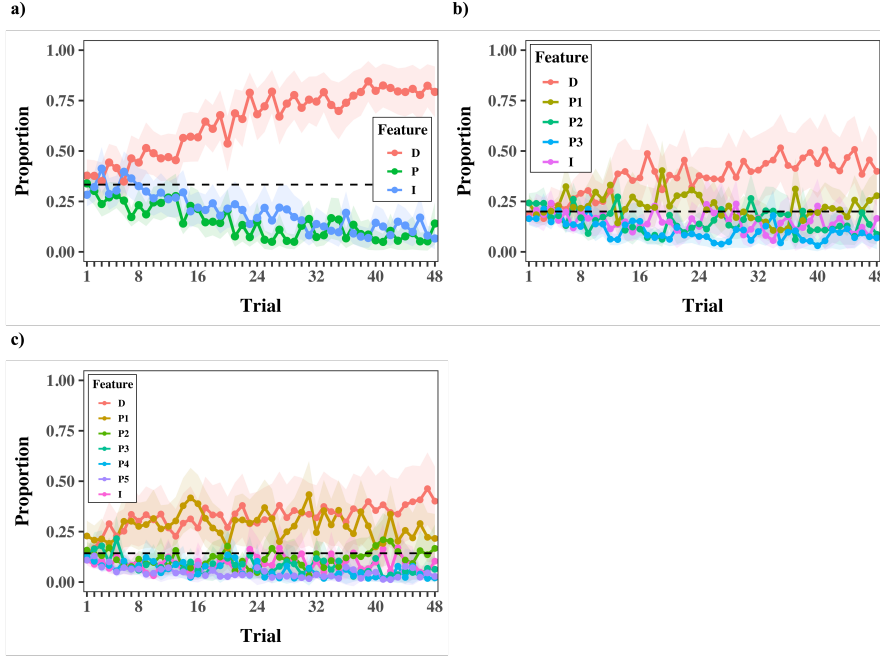


Fig. 2: Sampling trajectories for all deterministic, probabilistic, and irrelevant dimensions. This figure shows mean proportion of sampling time allocated to each dimension (color) across training in the three- (a), five- (b), and seven-dimensional (c) conditions, respectively. D denotes the deterministic feature, P denotes probabilistic features, and I denotes the irrelevant feature. In five- and seven-dimensional conditions, probabilistic features were rank-ordered separately for each participant according to total sampling time across training. P1 indicates the participant-specific most preferred probabilistic feature, P2 as the second most probabilistic feature, and so forth. Colored bands represent 90% confidence intervals based on the sampling distribution. Dashed black lines indicate chance-level sampling proportions of $1/3$, $1/5$, and $1/7$ for the three-, five-, and seven-dimensional conditions, respectively.

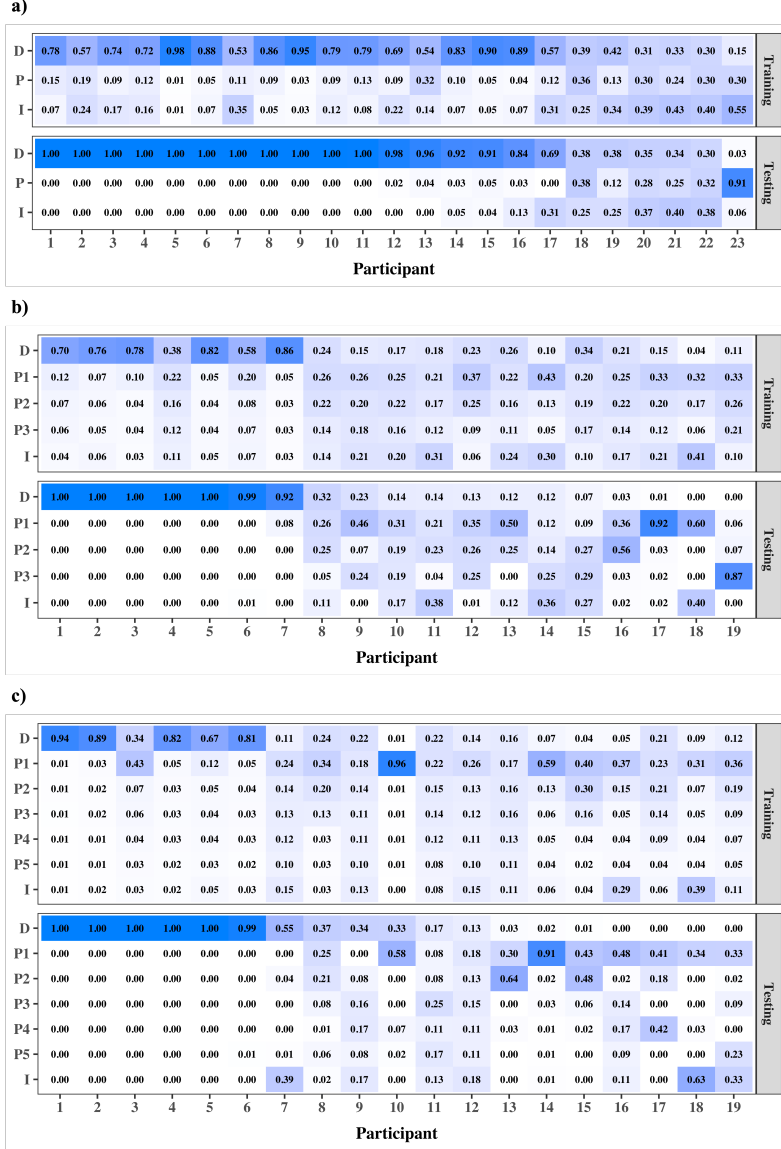


Fig. 3: Individual sampling proportions during training and test phases. Heatmaps show sampling profiles for each participant in the three- (a), five- (b), and seven-dimensional (c) conditions. For each condition, the upper panel summarizes the sampling during the training trials and the lower panel summarizes testing trials. Each column represents a participant's data and each row represents the data for a feature. Cell values and color intensity indicate the mean proportion of sampling time allocated to each feature, computed by averaging trial-wise sampling proportions within each participant. Participants are ordered by their sampling proportion for the deterministic feature (D) during testing. In the five- and seven-dimensional conditions, probabilistic features were rank-ordered within each participant according to total sampling time during training.

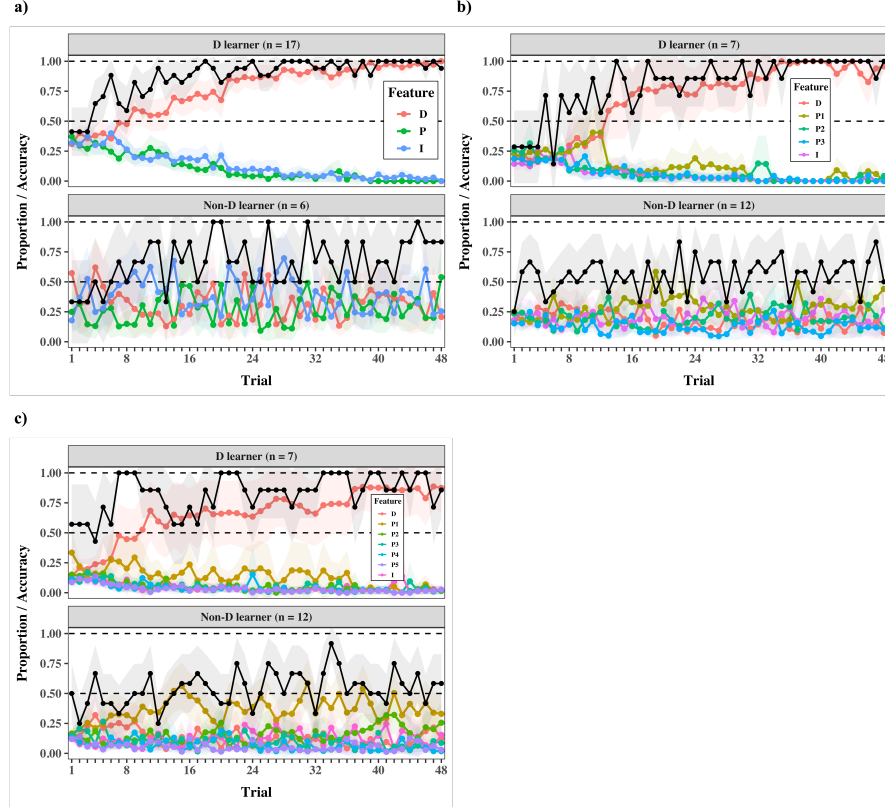


Fig. 4: Accuracy and feature-sampling patterns of D learners and non-D learners over learning. Training trajectories were shown separately for the three- (a), five- (b), and seven-dimensional (c) conditions, respectively. Within each condition, the upper panel shows D learners and the lower panel shows non-D learners, where D learners were defined as participants who allocated at least 50% of their sampling time during testing to the deterministic feature. Colored lines represent the mean proportion of sampling time allocated to each feature (color), with colored bands indicating 90% confidence intervals based on the sampling distribution. Black lines represent the mean accuracy, with gray bands indicating the corresponding 90% confidence intervals. Dashed lines indicate chance-level accuracy (.50) and perfect accuracy (1.00). In five- and seven-dimensional conditions, probabilistic features were rank-ordered within each participant according to total sampling time during training.

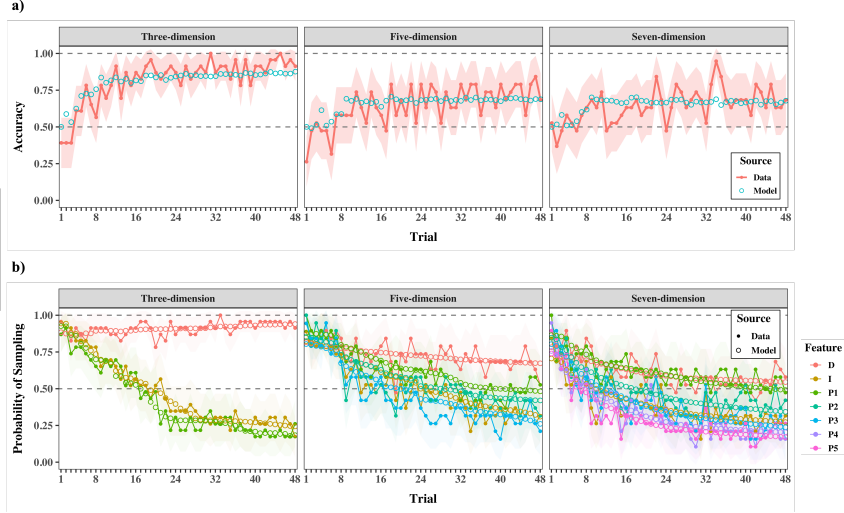


Fig. 5: Observed and error-effort model-predicted accuracy and sampling trajectories for all features. Observed and model-predicted accuracy trajectories are shown in the upper row (a), and observed and model-predicted feature-sampling trajectories are shown in the lower row (b), separately for the three- (left panel), five- (middle panel), and seven-dimensional (right panel) conditions. Filled dots and colored bands show observed group-level trajectories and 90% confidence intervals based on the sampling distribution. Open dots show trajectories simulated from the error-effort model with the best-fit parameters. Dashed horizontal lines mark chance-level accuracy (.50), perfect accuracy (1.00), and chance-level sampling probability (.50). In five- and seven-dimensional conditions, probabilistic features are displayed after participant-level rank ordering based on total sampling time during training.

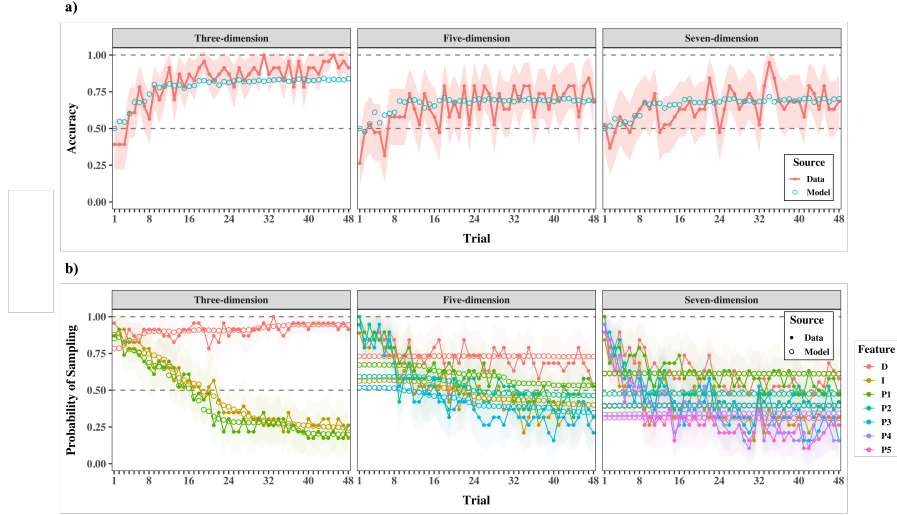


Fig. 6: Observed and ridge-regularization model-predicted accuracy and sampling trajectories for all features. Observed and model-predicted accuracy trajectories are shown in the upper row (a), and observed and model-predicted feature-sampling trajectories are shown in the lower row (b), separately for the three- (left panel), five- (middle panel), and seven-dimensional (right panel) conditions. The plotting conventions are the same as in Supplementary Figure 5, except that open dots show trajectories simulated from the ridge-regularization model with the best-fit parameters.

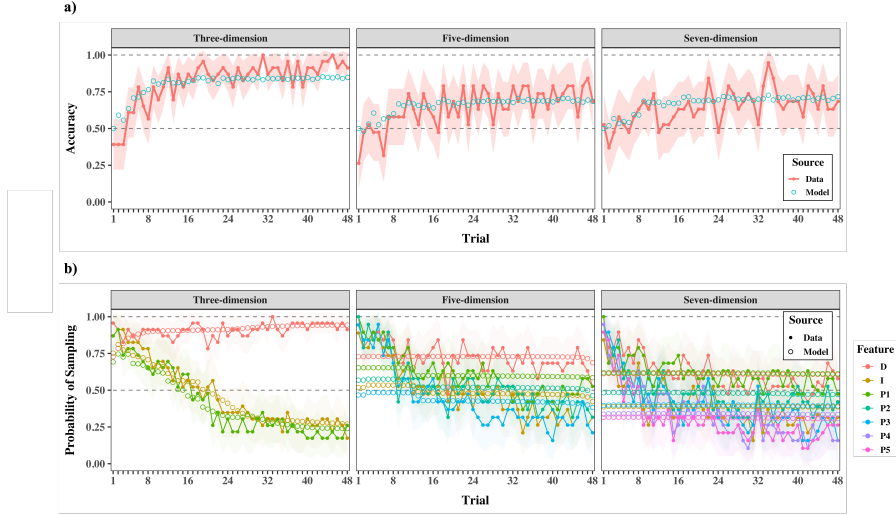


Fig. 7: Observed and ridge-regularization model-predicted accuracy and sampling trajectories for all features. Observed and model-predicted accuracy trajectories are shown in the upper row (a), and observed and model-predicted feature-sampling trajectories are shown in the lower row (b), separately for the three- (left panel), five- (middle panel), and seven-dimensional (right panel) conditions. The plotting conventions are the same as in Supplementary Figure 5, except that open dots show trajectories simulated from the lasso-regularization model with the best-fit parameters.

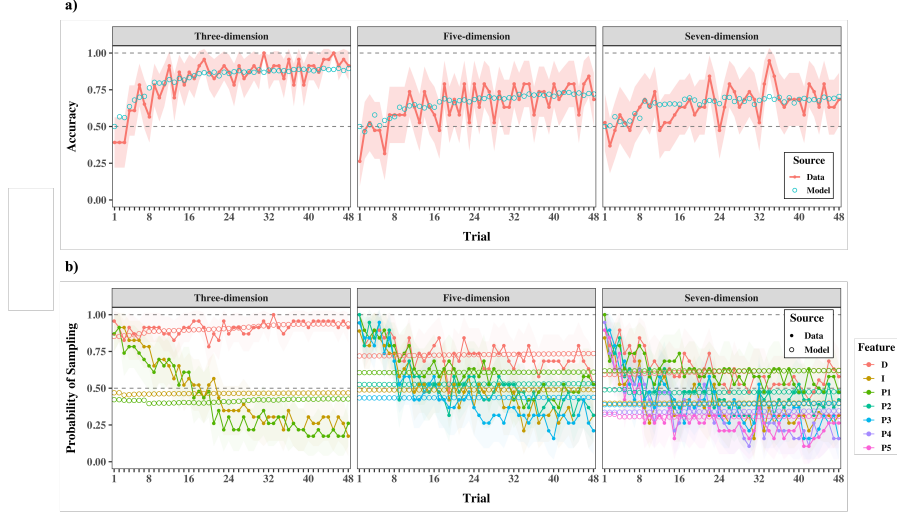


Fig. 8: Observed and ridge-regularization model-predicted accuracy and sampling trajectories for all features. Observed and model-predicted accuracy trajectories are shown in the upper row (a), and observed and model-predicted feature-sampling trajectories are shown in the lower row (b), separately for the three- (left panel), five- (middle panel), and seven-dimensional (right panel) conditions. The plotting conventions are the same as in Supplementary Figure 5, except that open dots show trajectories simulated from the unconstrained model with the best-fit parameters.

References

- Galdo, M., Weichart, E.R., Sloutsky, V.M., Turner, B.M. (2022, November). The quest for simplicity in human learning: Identifying the constraints on attention. *Cognitive Psychology*, 138, 101508, <https://doi.org/10.1016/j.cogpsych.2022.101508>