Supplemental Table S1. Estimated accuracy[1] of observed behaviors in zoo cheetahs provided different enrichment items.

Breakdown	% Agreement[2]	No. of Observations	Chi-Square	P-value
Overall	90.26	195	126.40	<0.001
Behavior				
Engagement	84.62	91	43.62	<0.001
Autogrooming	94.01	50	38.72	<0.001
Allogrooming	94.74	38	30.42	<0.001
No Behavior	100.00	16	.	.
Item				
Cinnamon	93.94	33	25.48	<0.001
Cologne	90.91	22	14.73	<0.001
Firehose log	96.00	25	21.16	<0.001
Jolly Egg	92.11	38	26.95	<0.001
Mirror	86.00	50	25.92	<0.001
Movie	85.19	27	13.37	<0.001
Category				
Olfactory	92.73	55	40.16	<0.001
Tactile	93.65	63	48.01	<0.001
Visual	85.71	77	39.28	<0.001
Time of Day				
AM	92.86	70	51.43	<0.001
PM	88.80	125	75.27	<0.001
Replicate				
1st Daytime	88.57	35	20.82	<0.001
1st Overnight	86.57	67	35.83	<0.001
2nd Daytime	97.14	35	31.11	<0.001
2nd Overnight	91.38	58	39.72	<0.001
Assessor				
A	84.21	38	17.78	<0.001
B	85.19	27	13.37	<0.001
C	100.00	7	.	.
D	85.29	34	16.94	<0.001
E	90.91	11	7.36	0.006
F	90.92	22	14.73	<0.001
G	100.00	42	.	.
H	92.86	14	10.28	<0.001

Reason for Disagreement[3]	% of Disagreements	% of Total Observations
Disparity in event timing/duration	15.8	1.7
Miss-identified behavior	26.3	2.8
Alternative Explanation for Behavior	42.1	4.5
Obvious Logging Typo/Data Entry Error	15.8	1.7

[1] Chi-squared tests were performed on a randomly selected subset of observations to determine the frequency at which the two assessors were in agreement regarding the observation. The subset included 15% of the observations logged by each assessor and two randomly-selected 5-minute periods/installment for which no events were logged.
[2] Agreement required identifying the same type of behavior and a similar duration for the behavioral event (± 20%).
[3] Disagreements were characterized into four categories. Disparity in event timing was defined as correctly-identified behaviors that were logged at substantially different durations (±20%). Misidentified behaviors were when one behavior was incorrectly identified as another (e.g., allogrooming rather than simultaneous autogrooming), as determined by the behavioral expert. Alternative explanations were when an apparent engagement behavior was incorrect because it was most likely due to circumstances not involving the enrichment item (e.g., cat appearing to briefly stare at the mirror was most likely watching a passing zoo employee), as determined by the expert. Obvious logging typos/data entry errors were when the evidence clearly indicated that a typo was the most likely explanation for the discrepancy (e.g., an event clearly ending at 11:01:03 was logged as ending at 11:10:03).

Supplemental Table S2. Non-parametric ranked sums analyses[1] for time spent performing behaviors in zoo cheetahs provided different environmental enrichment items.

Item	Average Time Spent in Sec/Hour			Ranked Sums of Average Time Spent		
	Enrichment			Enrichment		
	Sec/Hr	SE	*P*	Rank	SE	*P*
Firehose Log	15.4	5.9	a	12.0	2.4	a
Jolly Eggs	12.4	4.1	a	10.4	2.1	a
Mirror	292.4	194.1	b	20.8	0.9	b
Movie	404.1	277.2	b	18.0	3.6	b
Cinnamon	13.3	6.5	a	9.9	2.8	a
Cologne	2.3	1.7	c	4.0	2.0	c
P-values						
ENRICHMENT:			0.04			<0.01

	W	*P*
Shapiro-Wilk Test	0.466	<0.01

NS, not significant; SE, standard error.

[1] Data were analyzed by ANOVA on raw data (left side, presented in manuscript) and ranked sums of the data (right side) using the mixed procedure of SAS (SAS Institute, Cary, NC) to determine the effect of enrichment item, with rep as a random variable. Each item was provided four times (twice each during day and night), and replication was used as the experimental unit (n = 4). Shapiro-Wilk tests were performed to assess normality. Data are presented as mean ± standard error of the mean.

[a,b,c] Means with different superscripts differ ($P < 0.05$).

Supplemental Table S3. Non-parametric ranked sums analyses[1] for average duration of behavioral observations in cheetahs provided different environmental enrichment items.

| | | Average Length in Seconds | | | | | | | | | Ranked Sums of Average Length | | | | | | | | |
| | | Engaging | | | | | | | | | Engaging | | | | | | | | |
Enrichment	Time of Day	Enrichment	SE	P	Autogrooming	SE	P	Allogrooming	SE	P	Enrichment	SE	P	Autogrooming	SE	P	Allogrooming	SE	P
Firehose Log		.	.	.	41.5	5.2	.	.	.	.	.	.	.	184.5	10.2	.	.	.	.
Jolly Eggs		.	.	.	43.8	4.9	.	.	.	.	.	.	.	167.4	9.3	.	.	.	.
Mirror		.	.	.	35.1	9.3	.	.	.	.	.	.	.	180.3	19.9	.	.	.	.
Movie		.	.	.	99.7	29.6	.	.	.	.	.	.	.	222.8	23.6	.	.	.	.
Cinnamon		.	.	.	55.5	11.5	.	.	.	.	.	.	.	188.6	19.0	.	.	.	.
Cologne		.	.	.	52.6	9.6	.	.	.	.	.	.	.	204.4	11.9	.	.	.	.
	Day	.	.	.	30.9	7.8	a	.	.	.	.	.	.	154.7	9.7	a	.	.	.
	Night	.	.	.	78.6	8.8	b	.	.	.	.	.	.	227.9	9.5	b	.	.	.
Firehose Log	Day	12.7	4.1	a	19.6	7.8		14.8	9.1	x	175.0	22.4	a	128.9	15.2		87.2	8.8	a
Jolly Eggs	Day	5.5	4.3	ab	16.3	6.1		23.6	12.3	xy	88.3	20.5	b	116.0	11.6		91.4	12.1	ab
Mirror	Day	33.0	4.3	c	16.0	17.1		46.1	23.9	xy*	238.3	12.3	c	137.9	36.8		166.5	19.1	c
Movie	Day	47.2	7.7	d	68.6	38.2		19.6	37.0	xy	242.4	13.0	c	185.3	30.5		107.3	23.5	ad
Cinnamon	Day	2.1	4.0	b	30.2	13.5		32.6	10.8	y	29.6	23.4	d	159.6	22.3		110.6	13.2	bd
Cologne	Day	None	.		34.6	11.7		36.0	12.0	y	None	.		200.74	14.6		130.8	16.6	cd
Firehose Log	Night	42.7	5.6	a	63.5	6.9		99.9	11.4	x	324.61	30.8	a	240.07	13.4		207.0	11.0	a
Jolly Eggs	Night	25.3	5.2	b	71.3	7.7		82.8	14.2	xy	204.94	25.1	b	218.76	14.4		171.4	14.0	b
Mirror	Night	21.7	3.6	b	54.3	7.0		77.7	13.8	xy*	204.77	10.4	b	222.64	15.0		177.0	11.0	b
Movie	Night	61.7	25.6	a	130.8	45.3		258.5	58.5	z	303.44	43.3	a	260.25	36.1		260.3	37.2	a
Cinnamon	Night	23.0	2.8	b	80.9	18.7		83.1	16.0	xy	240.66	16.2	b	217.54	30.9		163.6	19.4	b
Cologne	Night	14.9	3.0	b	70.6	15.1		62.7	11.0	y	218.3	23.1	b	208.2	18.9		154.0	15.3	b

P-values

ENRICHMENT:				.			NS			NS			.			NS			.
TIME OF DAY:				.			<0.01			<0.01			.			0.03			.
ENR x TOD:				<0.01			NS			0.08			<.01			NS			0.01

Shapiro-Wilk		W	P		W	P		W	P										
		0.533	<0.01		0.598	<0.01		0.654	<0.01										

NS, not significant; SE, standard error.

[1] Data were analyzed by ANOVA on raw data (left side, presented in manuscript) and ranked sums of the data (right side) using the mixed procedure of SAS (SAS Institute, Cary, NC) to determine the effect of enrichment item, time of day, and the interaction, with rep as a random variable. Each item was provided four times (twice each during day and night), and replication was used as the experimental unit (n = 4). Shapiro-Wilk tests were performed to assess normality. Data are presented as mean ± standard error of the mean.

[a,b,c,d] Means with different superscripts differ ($P < 0.05$).

[x,y] Means with different superscripts tend to differ ($P < 0.10$).