

Table S1 Preregistration Deviations Table (adapted from Willroth &amp; Atherton, 2024)

Preregistration deviations						
#	Details		Original wording	Deviation description	To what extent is this a deviation from the preregistered plan?	Judgment of impact
1	Experiment	1, 2a, 2b	Our other dependent variables include $d'$ and $C$ . $d'$ and $C$ in each trip will be calculated with hit rate (ratio of items on the list selected) and false alarm (ratio of selected items that were not on the list), then averaged across trials for each participant	Instead of using $d'$ and $C$ as the index for performance, we opted for Streak Correct which was a measure used by similar studies.	<i>Major</i> -	<p>This deviation has a small impact on the results.</p> <p>The deviation was due to the high number of trips without errors (~90%), in which case <math>d'</math> and <math>C</math> could not be properly calculated.</p> <p>The risk of bias is low because we provided the results for <math>d'</math> and <math>C</math> in the Supplemental Materials, and they showed the same trend as the result using the Streak Correct measure.</p>
	Type	Anal... -				
	Reason	New ... -				
	Timing	After ... -				
2	Experiment	1, 2a, 2b	Our main dependent variable is dwell time. Mean dwell time (the average dwell time per trip) will be calculated for each trial, then averaged for each participant... Our other dependent variables include $d'$ and $C$ . $d'$ and $C$ in each trip will be calculated with hit rate (ratio of items on the list selected) and false alarm (ratio of selected items that were not on the list), then averaged across trials for each participant.	We preregistered to calculate and report all the measures, except Trips, by averaging across all trips. However, we focussed our discussion of the results from the first trip of each trial in the paper.	<i>Minor</i> -	<p>This deviation has a small impact on the results. The risk of bias is low.</p> <p>The results from all the trips and only the first trips were essentially the same. Therefore, we decided to mainly focus on the results from the first trips for a fair comparison with the One Shot condition in Experiment 2a/2b and to simplify the Results section. We have provided all the results from all the trips in the Supplemental Materials.</p>
	Type	Anal... -				
	Reason	New ... -				
	Timing	After ... -				

3	Experiment	1, 2a, 2b	<p>Exp1: Our main dependent variable is dwell time. Mean dwell time (the average dwell time per trip) will be calculated for each trial, then averaged for each participant...</p> <p>Exp2a &amp; 2b: Mean dwell time per trip will be calculated, then averaged within each participant...</p>	We calculated the dependent variable by averaging the data across trips for each trial for each participant, meaning that each observation was for a trial instead of a participant	<i>Minor</i> ▾	<p>This deviation has a small impact on the results. The risk of bias is low.</p> <p>We planned to analyze the data with the trial data. The description in the preregistration was a mistake in our wording.</p>
	Type	Anal... ▾				
	Reason	Typo/... ▾				
	Timing	Befor... ▾				
4	Experiment	1	<p>Additional variables: We will collect demographic data from children's parents, including the educational level (school grade) of the child, typical weekly screen time of the parents, typical weekly screen time of the child. These data will be used in exploratory analyses.</p>	We collected these data but did not analyze it.	<i>Minor</i> ▾	<p>This deviation has a small impact on the results.</p> <p>During data collection, we realized our questionnaire was not well-designed. In addition, since this analysis was preregistered as exploratory and was not central to the main questions of the paper, we decided not to analyze the data.</p> <p>Since this deviation was decided before analysis, the risk of bias is low.</p>
	Type	Anal... ▾				
	Reason	New ... ▾				
	Timing	Durin... ▾				

5	Experiment	2a, 2b	Dwell times that are longer than 30 seconds will be excluded... Participants must have completed all three trials in each condition to be included in the final analysis.	Instead of excluding trials that with Dwell Times > 30 s, we winsorized them by setting all Dwell Times > 30 s to 31 s	<i>Major</i> ▾	<p>This deviation has a small impact on the results.</p> <p>We planned the 30 s exclusion criterion based on data from Exp. 1 without anticipating that some children might dwell longer in the new, One-Shot condition. If we kept this criterion, we would have lost meaningful data. Therefore, instead of excluding them, we decided to winsorize these large values.</p> <p>We realized this issue when we were pre-processing the data. The risk of bias is low.</p> <p>In addition, results using the original exclusion are shown in the Supplemental Materials (Table S11-12); these show a similar pattern as the results reported in the main text.</p>
	Type	Anal... ▾				
	Reason	New ... ▾				
	Timing	After ... ▾				

**Table S2. Descriptive statistics across all trips**

	N	Trips		Dwell Time <sup>a</sup>		d' <sup>b</sup>		C <sup>b</sup>		Streak Correct <sup>c</sup>	
		M	SD	M	SD	M	SD	M	SD	M	SD
Exp1											
• No-Delay	34	4.07	1.73	4.57	2.47	1.08	0.23	1.19	0.15	1.87	0.96
• Long-Delay	34	3.52	1.73	5.86	3.21	1.16	0.27	1.14	0.17	2.22	1.15
Exp 2a											
• No-Delay	72	3.88	1.40	6.15	4.32	1.07	0.20	1.19	0.11	1.83	0.71
• Long-Delay	72	3.32	1.14	7.46	3.55	1.11	0.21	1.14	0.11	2.03	0.60
Exp 2b											
• No-Delay	35	4.69	1.53	3.84	1.90	0.99	0.19	1.26	0.11	1.48	0.56
• Long-Delay	35	3.59	1.35	5.10	1.63	1.10	0.24	1.17	0.11	1.95	0.68

Note: [All measures were averaged across all trips (as opposed to just first trips; see main text) for each trial. We included measures that were reported in the main text for convenience of comparison.]

<sup>a</sup> [We had preregistered to exclude trials that had Dwell Times > 30 s based on pilot data and Exp. 1 without anticipating that children might dwell longer in the One-Shot condition of Exp. 2a and 2b. In Exp 2a, ~6% of trials had a dwell time of over 30 s. Of these, 26 were in the One-Shot condition. 18 participants had trials with >30 s, and 12 were in the One-Shot condition. Given this, instead of excluding these values, we winsorized them by setting all Dwell Times > 30 s to 31 s, for all experiments (Shete et al., 2004). See Tables S8/S9 for results based on the original, strict exclusion.]

<sup>b</sup> [d' and C measures were pre-registered, but were unable to be calculated due to too few false alarms (see main text). They are included here for completeness. We calculated the hit rate (H, number of correctly picked items / total number of remaining correct items) and false alarm rate (FA, number of incorrectly picked items / total number of remaining incorrect items) for each trip. Because of the low FA rate (94% of trips, participants did not pick any incorrect items), we used the loglinear correction (Stanislaw & Todorov, 1999):  $H^* = \text{number of correct picks} + 0.5 / \text{number of correct items in a trip} + 1$ ,  $FA^* = \text{number of incorrect picks} + 0.5 / \text{number of incorrect items in a trip} + 1$ . (The number of correct and incorrect items changes from trip to trip.) d' was calculated as  $z(H^*) - z(FA^*)$  and C was calculated as  $[z(H^*) + z(FA^*)]/2$ . (z indicates Z-score.)]

<sup>c</sup> [Since false alarm rates were almost always zero, this forced a 'correction' for nearly all calculations of d' and C. Such corrections are meant to be applied only rarely (Brown & White, 2005), so it was no longer valid to proceed with these planned analyses. Instead, we turned to a straightforward measure used in similar work, Streak Correct, i.e. the length of the run of correct choices that precedes the first incorrect choice (or end-of-trial).]

**Table S3. Descriptive statistics of the first trip**

		Dwell Time		$d'$		$C$		Streak Correct	
	$N$	$M$	$SD$	$M$	$SD$	$M$	$SD$	$M$	$SD$
Exp1									
• No-Delay	34	5.81	3.70	0.84	0.31	1.21	0.19	2.00	1.11
• Long-Delay	34	6.77	4.00	0.94	0.40	1.14	0.21	2.42	1.32
Exp 2a									
• No-Delay	72	8.18	6.01	0.83	0.29	1.21	0.15	1.97	0.90
• Long-Delay	72	8.76	5.13	0.84	0.36	1.15	0.18	2.18	0.88
• One-Shot	72	14.36	7.86	0.95	0.47	0.83	0.24	3.23	1.27
Exp 2b									
• No-Delay	35	4.54	3.32	0.73	0.31	1.30	0.14	1.55	0.77
• Long-Delay	35	5.74	2.37	0.89	0.34	1.19	0.14	2.16	0.91
• One-Shot	35	9.46	6.49	0.77	0.44	0.89	0.25	2.56	0.82

Note: [We only included the first trip of each trial. We included measures here that were reported in the main text for convenience of comparison.]

**Table S4. Analyses across all trips (comparing No-Delay and Long-Delay conditions)**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$
Exp 1	4.50	.034	0.36	15.61	< .001	0.69	5.67	.017	0.41	9.57	.002	0.54	9.56	.002	0.54
Exp 2a	5.95	.015	0.29	50.06	< .001	0.84	2.22	.136	0.18	24.31	<.001	0.59	16.39	<.001	0.48
Exp 2b	10.81	.001	0.56	28.66	< .001	0.92	15.26	<.001	0.67	28.20	<.001	0.91	30.84	<.001	0.95

Note: [All measures, except for Trips, were averaged across all the trips for each trial. We included measures that were reported in the main text for convenience of comparison.]

Table S5. Results of the analyses of the first trip with pairwise post-hoc tests comparing all three conditions

	Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/z$	$p$	$d/\eta_p^2$
<b>Exp 1</b>	<b>7.96</b>	<b>.005</b>	<b>0.49</b>	<b>3.35</b>	<b>.067</b>	<b>0.32</b>	<b>6.68</b>	<b>.010</b>	<b>0.45</b>	<b>3.41</b>	<b>.065</b>	<b>0.32</b>
<b>Exp 2a</b>	<b>79.62</b>	<b>&lt; .001</b>	<b>0.53</b>	<b>22.83</b>	<b>&lt; .001</b>	<b>0.16</b>	<b>162.90</b>	<b>&lt;.001</b>	<b>0.70</b>	<b>65.97</b>	<b>&lt;.001</b>	<b>0.36</b>
• Long-Delay vs.No-Delay	3.36	.001	0.40	0.90	.370	0.11	-3.72	< .001	0.44	1.75	.081	0.21
• Long-Delay vs.One-Shot	-7.27	< .001	0.86	-3.91	< .001	0.46	10.10	< .001	1.20	-5.43	< .001	0.64
• No-Delay vs.One-Shot	-8.88	< .001	1.05	-4..69	< .001	0.56	12.75	< .001	1.51	-7.65	< .001	0.90
<b>Exp 2b</b>	<b>37.54</b>	<b>&lt; .001</b>	<b>0.52</b>	<b>9.20</b>	<b>.010</b>	<b>0.14</b>	<b>136.00</b>	<b>&lt;.001</b>	<b>0.80</b>	<b>24.60</b>	<b>&lt;.001</b>	<b>0.43</b>
• Long-Delay vs.No-Delay	4.34	< .001	0.74	3.01	.015	0.52	-3.72	< .001	0.64	3.16	.003	0.53
• Long-Delay vs.One-Shot	-3.66	< .001	0.63	1.58	.247	0.27	6.88	< .001	1.18	-1.81	.070	0.31
• No-Delay vs.One-Shot	-6.03	< .001	1.03	-0.60	.549	-0.10	11.02	< .001	1.89	-4.96	< .001	0.84

Note: [We only included the first trip of each trial. The main effect of the condition is in bold. For comparison between two conditions, we report  $z$  or  $t$  and Cohen's  $d$ . For comparison among three conditions, we report  $\chi^2$  and partial  $\eta^2$ . We included results that were reported in the main text for the convenience of comparison among measures.]

**Table S6. The main effect of age and the interaction between condition (No-Delay and Long-Delay) and age in Exp. 1**

	Trips			Dwell Time			$d'$			C			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
Across all trips															
• Age main effect	5.44	.020	0.15	1.47	.225	0.01	5.72	.017	0.15	4.09	.043	0.11	5.28	.022	0.14
• Interaction effect	2.18	.139	0.06	10.50	.001	0.25	5.96	.015	0.16	1.24	.264	0.04	3.36	.067	0.10
Only first trip															
• Age main effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.38	.537	0.004	1.93	.165	0.08	2.80	.094	0.07	2.77	.096	0.08
• Interaction effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	2.26	.133	0.07	1.28	.259	0.01	0.25	.612	0.002	0.03	.860	0.0003

*Note: [Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition)]*



**Table S7. The main effect of trial number and the interaction between condition and trial number (learning effect) in Exp. 1**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
Across all trips															
• Trial main effect	0.05	.825	0.0004	0.12	.731	0.0009	1.98	.160	0.01	0.17	.681	0.001	0.17	.684	0.001
• Interaction effect	0.01	.907	0.0001	0.19	.661	0.001	0.15	.696	0.001	0.01	.915	0.0001	0.04	.848	0.0003
Only first trip															
• Trial main effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.60	.440	0.004	0.95	.329	0.01	1.31	.253	0.01	0.003	.954	0.00002
• Interaction effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.07	.785	0.0006	0.74	.391	0.004	0.03	.863	0.0002	0.17	.678	0.001

*Note: [Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition)]*

Table S8. The interaction between condition and Metacognition in Exp. 2a and Exp. 2b

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
<b>Exp 2a</b>															
Easier															
• Across all trips	0.42	.519	0.006	0.01	.942	0.0001	0.89	.346	0.01	0.77	.379	0.01	1.41	.235	0.02
• Only first trip	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.26	.613	0.004	3.24	.072	0.04	2.22	.136	0.03	1.65	.199	0.02
Preference															
• Across all trips	0.09	.763	0.001	0.51	.473	0.007	0.00001	.997	0.0000002	0.86	.355	0.01	0.07	.796	0.001
• Only first trip	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.03	.873	0.0004	0.02	.895	0.0003	3.56	.059	0.05	0.21	.646	0.003
<b>Exp 2b</b>															
Easier															
• Across all trips	0.43	.512	0.01	0.22	.636	0.007	0.21	.649	0.006	0.63	.426	0.02	0.56	.453	0.02
• Only first trip	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.61	.435	0.02	0.05	.821	0.002	0.22	.643	0.006	0.07	.788	0.002
Preference															
• Across all trips	0.04	.846	0.001	0.30	.586	0.01	0.18	.675	0.005	1.26	.262	0.04	0.18	.676	0.01
• Only first trip	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.48	.490	0.01	0.01	.917	0.0003	0.67	.414	0.02	0.21	.643	0.01

Note: [Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition). Easier question: Which one is easier? Preference question: Which one do you want to play again?]

**Table S9. The main effect of sites (China/USA) and the interaction between Conditions and Sites (Comparison between Exp. 2a and 2b)**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
Across all trips															
• Site main effect	3.45	.063	0.05	4.41	.036	0.07	0.79	.375	0.006	2.94	.086	0.05	1.63	.202	0.03
• Interaction effect	1.14	.286	0.02	0.94	.332	0.01	2.02	.155	0.03	2.32	.128	0.03	2.95	.086	0.04
Only first trip															
• Site main effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	3.88	.049	0.08	0.50	.482	0.02	3.23	.073	0.03	1.74	.187	0.03
• Interaction effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	3.53	.172	0.05	3.39	.184	0.03	1.80	.406	0.03	2.63	.268	0.04

*Note: [We only compared condition order B and D here. Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition)]*

**Table S10. The main effect of position and the interaction between condition and One-Shot block position (*preceding or following* {No-Delay/Long-Delay}) in Exp. 2a**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
Across all trips															
• Position main effect	1.14	.286	0.02	8.33	.004	0.10	0.34	.561	0.005	0.73	.392	0.01	0.45	.504	0.01
• Interaction effect	0.008	.927	0.0001	0.01	.914	0.0002	0.24	.621	0.003	0.007	.933	0.0001	0.18	.675	0.003
Only first trip															
• Position main effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	8.20	.004	0.10	1.31	.253	0.02	0.86	.354	0.01	2.11	.146	0.03
• Interaction effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.29	.866	0.004	0.63	.731	0.01	0.34	.844	0.005	0.37	.831	0.004

*Note: [Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition)]*

**Table S11. Descriptive statistics for Exp. 2a (excluding participants with Dwell Times > 30 s)**

	N <sup>a</sup>	Trips		Dwell Time		d'		C		Streak Correct	
		M	SD	M	SD	M	SD	M	SD	M	SD
Across all trips											
• No-Delay	66	3.98	1.35	5.19	2.39	1.05	0.18	1.21	0.09	1.73	0.55
• Long-Delay	66	3.36	1.17	6.76	2.55	1.10	0.21	1.14	0.12	2.00	0.60
Only first trip											
• No-Delay	55	n/a	n/a	6.92	4.31	0.77	0.25	1.23	0.14	1.82	0.73
• Long-Delay	55	n/a	n/a	7.57	3.37	0.77	0.37	1.16	0.18	2.03	0.84
• One-Shot	55	n/a	n/a	11.08	5.30	0.83	0.42	0.84	0.26	2.86	1.13

Note: [Here we only report the result of Exp. 2a. In Exp. 1, no participant had a trip that exceeded 30 s. In Exp. 2b, there were only two participants with >30 s dwell time (30.40 s and 30.97s), which were not flagged as outliers. Results 'Across all trips' are averaged across all trips and only relevant for the No-Delay and Long-Delay conditions (there was only one trip, by design, in the One-shot condition). For 'Only first trip' results, we included the first trip for each trial for all three conditions (No-Delay, Long-Delay, and One-Shot condition)]

<sup>a</sup> [When only considering the No-Delay/Long-Delay condition for the analyses of all the trips, 6 participants had trips with > 30 s dwell time and were excluded, resulting in 66 participants. When considering all three conditions for the analyses of the first trip, 17 participants had trips with > 30 s dwell time and were excluded, resulting in 55 participants. ]

**Table S12. Additional analyses for Exp. 2a (excluding participants with Dwell Times > 30 s)**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$	$\chi^2/t$	$p$	$d/\eta_p^2$
Across all trips	<b>7.48</b>	<b>.006</b>	<b>0.34</b>	<b>57.00</b>	<b>&lt;.001</b>	<b>0.94</b>	<b>4.77</b>	<b>.029</b>	<b>0.27</b>	<b>28.41</b>	<b>&lt;.001</b>	<b>0.66</b>	<b>23.99</b>	<b>&lt;.001</b>	<b>0.61</b>
Only first trip				<b>51.69</b>	<b>&lt;.001</b>	<b>0.49</b>	<b>11.08</b>	<b>.004</b>	<b>0.11</b>	<b>110.49</b>	<b>&lt;.001</b>	<b>0.67</b>	<b>39.86</b>	<b>&lt;.001</b>	<b>0.33</b>
• Long-Delay vs.No-Delay	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	2.94	.005	0.40	0.87	.388	0.12	-3.15	.003	0.43	1.19	.236	0.16
• Long-Delay vs.One-Shot	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	-5.36	<.001	0.73	-2.38	.041	0.32	8.10	<.001	1.10	-4.65	<.001	0.63
• No-Delay vs.One-Shot	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	-7.18	<.001	0.98	-3.33	.005	0.45	10.46	<.001	1.42	-5.80	<.001	0.78

Note: [Here we only report the main effect of the condition, to see the results of all the same analyses that were reported in the main text for Exp. 2a, please refer to the detailed result shared in the OSF folder.]

**Table S13. Descriptive statistics for the combined analysis of data from all 5-6-year-olds across all three experiments in the No-Delay and Long-Delay conditions**

	<i>N</i>	Trips		Dwell Time		<i>d'</i>		<i>C</i>		Streak Correct	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Across all trips											
• No-Delay	125	4.19	1.54	5.26	3.67	1.04	0.20	1.22	0.11	1.71	0.70
• Long-Delay	125	3.52	1.34	6.43	3.17	1.10	0.22	1.15	0.12	1.98	0.67
Only first trip											
• No-Delay	125	<i>n/a</i>	<i>n/a</i>	6.81	5.37	0.79	0.30	1.24	0.16	1.83	0.90
• Long-Delay	125	<i>n/a</i>	<i>n/a</i>	7.46	4.49	0.85	0.36	1.16	0.18	2.16	0.94

Note: [As suggested by a reviewer, we present results based on combined data from all 5-6-year-old children across all three experiments ( $N=125$ ). Here we only included 5-6-year-olds from Experiment 1 ( $N = 18$ ) to match the age range of children in Experiment 2a&2b. In addition, we only included data from No-Delay and Long Delay condition from Experiment 2a&2b because there were only these two conditions in Experiment 1.]

**Table S14. Analyses of combined data from all three experiments (comparing No-Delay and Long-Delay conditions)**

	Trips			Dwell Time			$d'$			$C$			Streak Correct		
	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$	$\chi^2$	$p$	$d$
Across all trips	11.88	<.001	0.31	80.30	<.001	0.80	10.86	<.001	0.30	50.45	<.001	0.64	42.64	<.001	0.59
Only first trip	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	34.42	<.001	0.53	3.61	.058	0.17	31.29	<.001	0.50	11.34	<.001	0.30

Note: [As suggested by a reviewer, we present results based on combined data from all 5-6-year-old children across all three experiments (N=125). Here we only included 5-6-year-olds from Experiment 1 (N = 18) to match the age range of children in Experiment 2a&2b. In addition, we only included data from No-Delay and Long Delay condition from Experiment 2a&2b because there were only these two conditions in Experiment 1.]

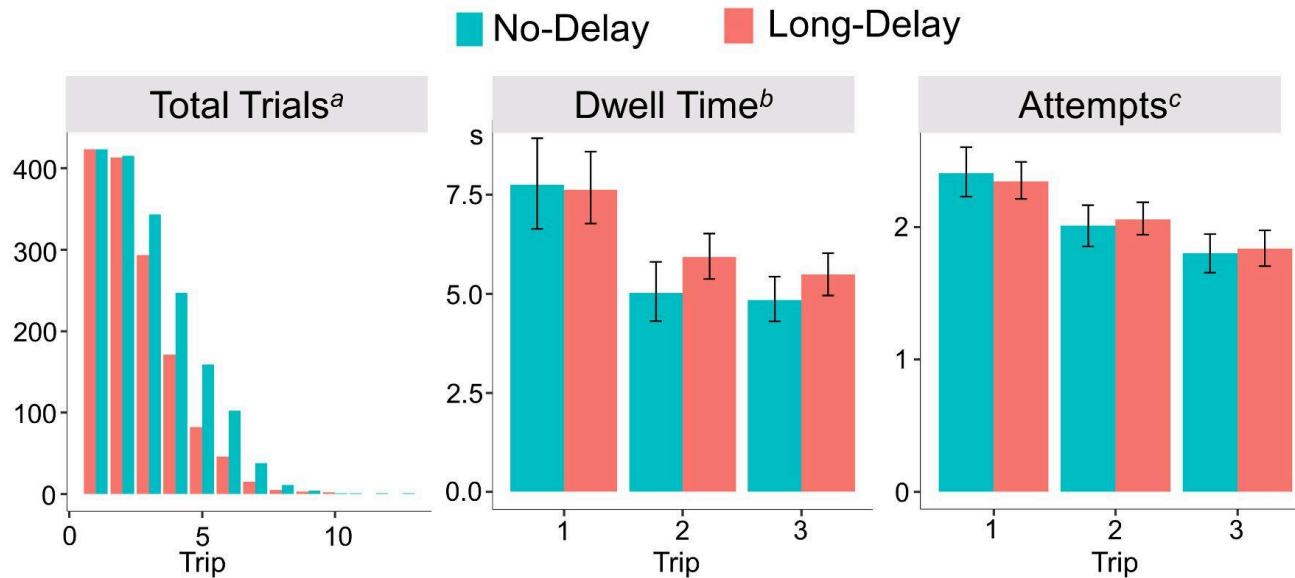


**Table S15. Analyses with combined data from all three experiments for order effect (No-Delay first or Long-Delay conditions first)**

	Trips			Dwell Time			$d'$			C			Streak Correct		
	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$	$\chi^2$	$p$	$\eta_p^2$
Across all trips															
• Order main effect	7.06	.008	0.05	0.24	.623	0.0005	1.92	.166	0.01	5.63	.018	0.04	3.11	.078	0.02
• Interaction effect	0.03	.869	0.0002	0.41	.523	0.003	0.50	.480	0.004	0.14	.712	0.001	0.001	.969	0.00001
Only first trip															
• Order main effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.12	.728	0.001	2.13	.145	0.009	6.58	.010	0.05	5.40	.020	0.04
• Interaction effect	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	0.005	.946	0.00004	2.21	.137	0.02	0.64	.422	0.005	0.44	.508	0.004

Note: [As suggested by a reviewer, we present results based on combined data from all 5-6-year-old children across all three experiments (N=125). Here we only included 5-6-year-olds from Experiment 1 (N = 18) to match the age range of children in Experiment 2a&2b. In addition, we only included data from No-Delay and Long Delay condition from Experiment 2a&2b because there were only these two conditions in Experiment 1.]

Figure S1. Combined data from all three experiments for trip effect with No-Delay and Long-Delay condition



Note: [As suggested by a reviewer, we added analyses to look for changes across trips and, more specifically, to gain insight into whether children were aiming for six items (the actual number of correct items a child needed to select to complete a trial; this was, by design, not explicitly revealed to the child) or 10 items (the length of the entire shopping list). As can be seen here in the first panel, nearly all trials contained at least two trips (~800, across both conditions), with a smaller set containing three or more (~600), and with the number of trials with yet longer trip lengths falling rapidly (see Total Trials, first panel). To have sufficient data to analyze trends, and to ensure more fair comparisons, the subsequent exploratory analyses here on Dwell Time and Attempts are based only on trials that finished in exactly 3 trips (N = 218 trials). (Note that this selection artificially minimizes the condition effect since, by explicitly equalizing the number of trips, it also tends to align the related measures of Dwell Time and Attempts.) Because of the potentially sudden end of the trial in our design (when the child accumulates six correct selections), we cannot evaluate whether children were aiming for six or 10 items from their number of attempts on their last trip. However, we can analyze the relative amount of effort children devoted to memorizing items from the list (Dwell Time) on their last trip, relative to preceding trips. Here, while there is an apparent drop from the first to the second trip, there is no significant decrease from the second to the third (last) trip. We take this as evidence that children did not alter their effort level for this last trip, in other words, they were likely prepared to continue making responses at the same effort level (out to 10 items, presumably) had the trial not ended.

<sup>a</sup> [Total Trials are the total number of trials with at least a given number of trips (similar to a survival analysis). For example, each child had at least one trip for each trial in both conditions, leading to a total trial count of 423 (141 children X 3 trials per condition) for the first trip. A few children were able to finish the whole trial in just one trip, especially in the Long-Delay condition, leading to a decreased total trial count for the second trips, 415 for the No-Delay condition, 413 for the Long-Delay condition, etc.]

<sup>b</sup> [Dwell Time in trials that finished in exactly 3 trips.]

<sup>c</sup> [Attempts per trip in trials that finished in exactly 3 trips. Attempts as a measure was suggested by a reviewer. This measure represents the number of items children picked per trip regardless of their accuracy (though, in the vast majority of the visits to the store, since children made so few errors (~10%), Attempts and Streak Correct are equivalent).]

## References

- Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, *37*(3), 436–449.
- Shete, S., Beasley, T. M., Etzel, C. J., Fernández, J. R., Chen, J., Allison, D. B., & Amos, C. I. (2004). Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection. *Behavior Genetics*, *34*(2), 153–159.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *31*(1), 137–149.
- Willroth, E. C., & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science*, *7*(1). <https://doi.org/10.1177/25152459231213802>