Comparison Axis alignment

Supplemental Material

1 Experiment Details

1.1 Generating Experimental Stimuli

The stimuli are generated by fully crossing the following five binary attributes:

- Longer bar side (2 levels: longer bar appears on left or right chart)
- Higher bar side (2 levels: higher bar appears on left or right chart)
- Later bar side (2 levels: bar with the higher index appears on left or right chart)
- Darker bar side (2 levels: darker bar appears on left or right chart)
- Difficulty (2 levels: easy or hard)

This results in $2^5 = 32$ unique stimuli combinations, and each combination is repeated 4 times to make up a total of 128 stimuli. This design ensures that the overall set of stimuli is balanced across key attributes while maintaining variety within each block.

The following constraints are applied to stimuli to control for confounds:

- The darkest bar of a chart never appears in the first two or the last two indexes.
- There are 10 bars.
- The range of the y-axis is 0 to 100.
- The length of bars ranges from 20 to 70.
- The "floor" (i.e., minimum) of each bar is at least 5.
- The "ceiling" (i.e., maximum) of each bar is at most 95.
- The color difference between the first and second darkest bars in each chart is fixed at 2 steps on a 8-step single-hue ordinal scale.
- The colors of the darkest bars in each chart are always different with the difference fixed at 1 step on a 8-step single-hue ordinal scale (for the engagement checks, the two darkest bars have the same color).
- The colors of the darkest bars in each chart are always one of the top 3 darkest colors on a 8-step single-hue ordinal scale.
- The lightest color in the 8-step color scale is not used.
- The height difference between the darkest bars across the two charts is fixed at 20 for easy difficulty, and 10 for hard difficulty (30 for engagement checks).
- The length difference between the darkest bars across the two charts is fixed at 20% for easy difficulty, and 10% for hard difficulty (40% for engagement checks).
- The index difference between the darkest bars across the two charts is fixed at 1 for easy difficulty, and 2 for hard difficulty (5 for engagement checks).
- The sum of all bar lengths in the two charts are equal.

For each stimulus (i.e., pair of charts), the height, length, and index of the darkest bars in each chart is first generated and colored randomly. The rest of the bars are then randomly generated and are colored randomly. The script for generating the stimuli can be found here: `https://osf.io/bjhrc`

1.2 Power Analysis

We used the results of the power analysis for a previous experiment, where we used the R mixedpower library to conduct a power analysis. In a prior experiment, we had 3 factors: task, layout, and labels. It was our goal to find a 3-way interaction between these three factors. We built a linear mixed-effects model based on this hypothesis and ran a power analysis with the goal of obtaining .90 power to detect a medium effect size. The results indiciated that 40 participants per task condition is needed.

1.3 Task Instructions

Figure 1: The general instruction that participants see first.

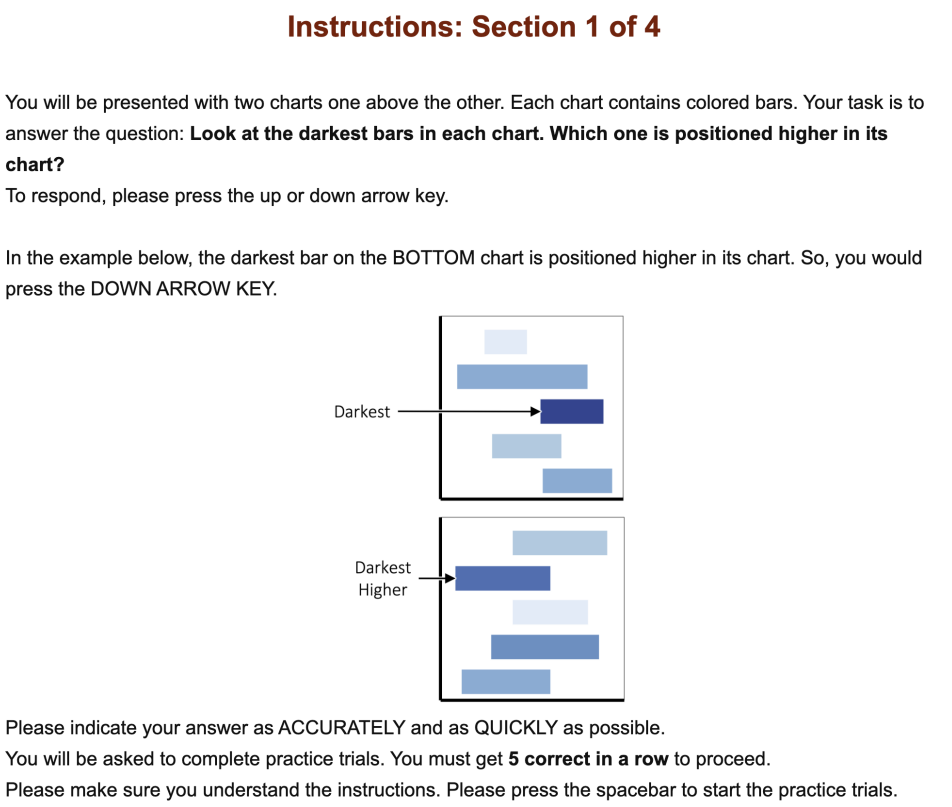

Figure 2: An example screenshot of an instruction page from the conditions index task - horizontal bars - stacked layout.

The instruction for the task looked as follows:

- (Stacked layout) “You will be presented with two charts one above the other.”
- (Side-by-side layout) “You will be presented with two charts side by side.”

“Each chart contains colored bars. Your task is to answer the question: Look at the darkest bars in each chart.”

- (Height task - vertical bars) “Which one reaches higher in its chart?”
- (Height task - horizontal bars) “Which one reaches farther right in its chart?”
- (Index task - vertical bars) “Which one is positioned farther right in its chart?”
- (Index task - horizontal bars) “Which one is positioned higher in its chart?”
- (Stacked layout) “To respond, please press the up or down arrow key.”
- (Side-by-side layout) “To respond, please press the left or right arrow key.”

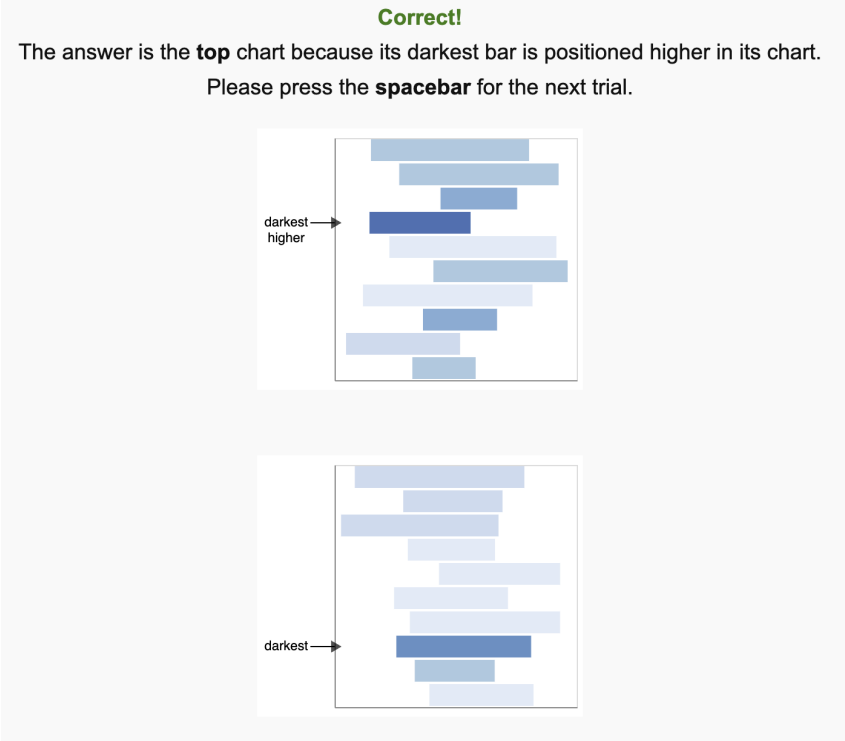

Figure 3: An example screenshot of an easy practice trial that explains the answer after each trial. It is taken from the conditions index task - horizontal bars - stacked layout.

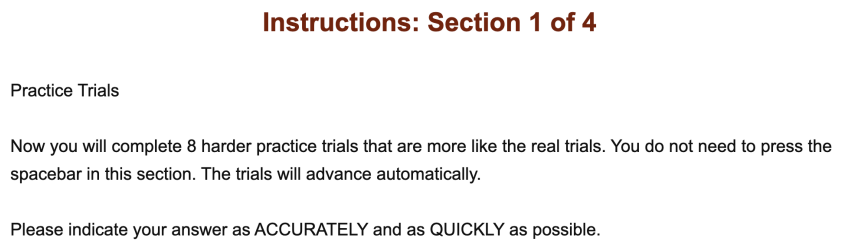

Figure 4: The instructions after the easy practice trials and before the hard practice trials.

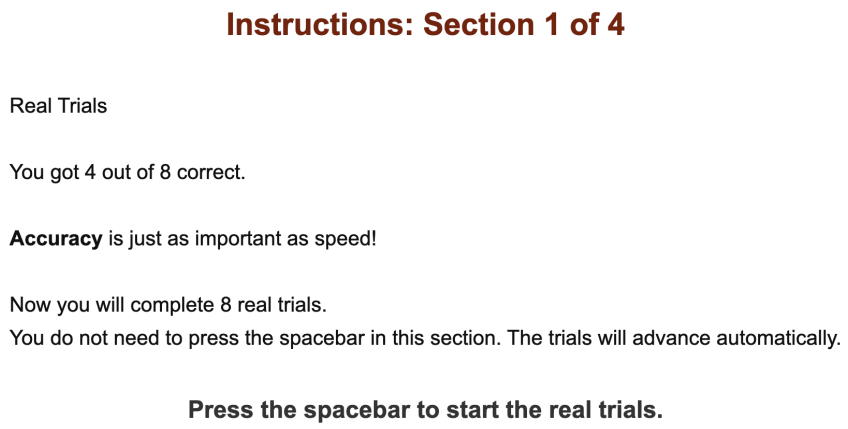

Figure 5: The instructions after the hard practice trials and before the real trials.

2 Hypotheses Analysis

Base model:

$$\text{RT} \sim \text{task}*\text{layout}*\text{orientation} + (1 + \text{layout}*\text{orientation} \mid \text{participant})$$

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1348.80	40.66	78.03	33.169	<2e-16 ***
task	-101.40	40.66	78.03	-2.493	0.01477 *
layout	42.19	14.06	77.36	3.001	0.00362 **
orientation	-20.74	11.50	78.13	-1.802	0.07534 .
task:layout	27.75	14.06	77.36	1.974	0.05193 .
task:orientation	26.94	11.50	78.13	2.342	0.02173 *
layout:orientation	15.17	12.26	77.61	1.237	0.21986
task:layout:orientation	-137.59	12.26	77.61	-11.219	<2e-16 ***

Table 1: Output of the base model.

Base model within each task:

$$\text{RT} \sim \text{layout}*\text{orientation} + (1 + \text{layout}*\text{orientation} \mid \text{participant})$$

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1247.165	53.375	39.015	23.366	<2e-16 ***
layout	70.542	18.261	38.197	3.863	0.000421 ***
orientation	6.671	13.905	39.051	0.480	0.634093
layout:orientation	-122.373	15.793	38.373	-7.749	2.29e-09 ***

Table 2: Output of the model within height task:

	Estimate	Std. Error	df	t-value	Pr(> |t|)
(Intercept)	1450.23	61.43	39.02	23.608	<2e-16 ***
layout	14.63	21.33	39.01	0.686	0.497
orientation	-47.51	18.26	39.06	-2.602	0.013 *
layout:orientation	152.74	18.77	39.09	8.135	6.14e-10 ***

Table 3: Output of the model within index task:

We then move on to analyzing this interaction in two ways: looking at the effect of layout within each task-bar orientation pair, and looking at the effect of bar orientation within each task-layout pair.

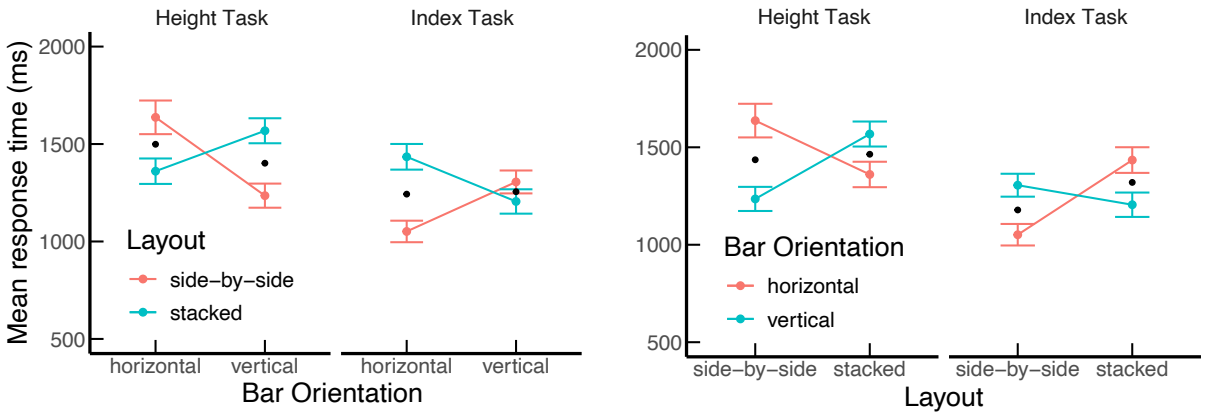

Figure 6: Mean response time grouped by task, layout, and bar orientation. Error bars represent the standard error of the mean. The left chart shows the layout on the x-axis and the bar orientation as the color. The right chart shows the bar orientation on the x-axis and the layout as the color.

2.1 Interaction analysis within each task and bar orientation

Base model within each pair of task and bar orientation:

$$\text{RT} \sim \text{layout} + (1 + \text{layout} \mid \text{participant})$$

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1498.47	69.74	38.99	21.488	<2e-16 ***
layout	-138.27	31.39	39.06	-4.405	7.99e-05 ***

Table 4: Output of the model within height task and horizontal bar conditions:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1402.04	57.88	39.11	24.222	<2e-16 ***
layout	167.26	25.14	38.92	6.654	6.49e-08 ***

Table 5: Output of the model within height task and vertical bar conditions:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1241.12	54.60	39.02	22.730	<2e-16 ***
layout	192.46	25.76	38.50	7.473	5.25e-09 ***

Table 6: Output of the model within index task and horizontal bar conditions:

	Estimate	Std. Error	df	t-value	Pr(>|t|)
(Intercept)	1254.42	55.82	39.00	22.475	<2e-16 ***
layout	-52.31	22.62	36.11	-2.313	0.0266 *

Table 7: Output of the model within index task and vertical bar conditions:

2.2 Interaction analysis within each task and layout

Base model within each pair of task and bar orientation:

$$\text{RT} \sim \text{orientation} + (1 + \text{orientation} \mid \text{participant})$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1436.01	69.26	38.99	20.73	<2e-16 ***
orientation	-200.37	28.91	38.94	-6.93	2.7e-08 ***

Table 8: Output of the model within height task and side-by-side layout conditions:

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1464.97	60.52	39.09	24.204	<2e-16 ***
orientation	105.20	23.16	39.00	4.542	5.25e-05 ***

Table 9: Output of the model within height task and stacked layout conditions:

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1176.78	52.97	38.99	22.215	<2e-16 ***
orientation	128.95	19.64	37.54	6.567	1.01e-07 ***

Table 10: Output of the model within index task and side-by-side layout conditions:

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1318.20	59.76	39.01	22.06	<2e-16 ***
orientation	-115.91	22.46	37.99	-5.16	8.05e-06 ***

Table 11: Output of the model within index task and stacked layout conditions:

2.3 Comparison axis alignment captures unexplained variance

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1348.66	42.01	79.05	32.105	<2e-16 ***
layout	42.60	14.34	78.28	2.971	0.00394 **
orientation	-20.35	11.81	78.85	-1.724	0.08870 .
alignment	-137.59	12.31	78.55	-11.177	<2e-16 ***

Table 12: Alignment captures independent variance in performance that is unexplained by layout or orientation alone.

3 Interference Analysis

We created extended models that add the fixed effect of factors that could potentially cause interference.

Model extended with answer side:

$$\text{RT} \sim \text{task}*\text{layout}*\text{orientation} + \text{answerSide} + (1 + \text{layout}*\text{orientation} + \text{answerSide} \mid \text{participant})$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1353.131	40.622	78.043	33.311	<2e-16 ***
task	-97.054	40.291	78.054	-2.409	0.0184 *
layout	43.551	14.049	77.382	3.100	0.0027 **
orientation	-20.198	11.575	78.128	-1.745	0.0849 .
answerSide	-74.409	8.326	79.488	-8.937	1.23e-13 ***
task:layout	26.850	14.048	77.404	1.911	0.0597 .
task:orientation	24.329	11.371	77.863	2.140	0.0355 *
layout:orientation	16.507	12.345	77.657	1.337	0.1851
task:layout:orientation	-139.338	12.228	77.569	-11.395	<2e-16 ***

Table 13: Outputs of model extended with answer side factor.

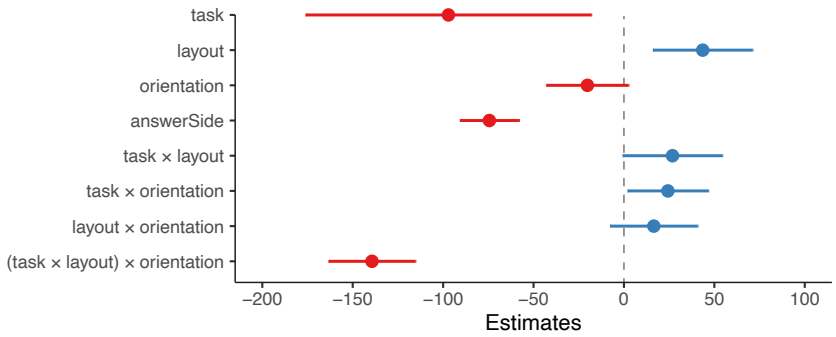

Figure 7: Coefficient estimate plot for model with fixed and random effect of answerSide. Error bars represent 95% confidence intervals. Negative coefficient indicates faster response time on tasks whose answer is the top chart or the right chart. 3-way interaction is still significant after accounting for the effect of interference.

We added four more factors as fixed effects to capture the interference caused by distractor tasks. Adding them as random effects caused the model to fail to converge, so we add them only as fixed effects.

Full extended model:

$$\begin{aligned} \text{RT} \sim\ & \text{task}*\text{layout}*\text{orientation} + \text{answerSide} + \text{isAnswerLonger} \\ & + \text{isAnswerDarker} + \text{isAnswerHigher} + \text{isAnswerLater} \\ & + (1 + \text{layout}*\text{orientation} + \text{answerSide} \mid \text{participant}) \end{aligned}$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1375.961	40.896	80.499	33.645	<2e-16 ***
task	-92.124	40.587	80.456	-2.270	0.025897 *
layout	43.521	14.133	77.321	3.079	0.002871 **
orientation	-19.525	11.607	78.074	-1.682	0.096523
answerSide	-75.546	8.274	79.501	-9.131	5.13e-14 ***
isAnswerLonger	-12.472	5.027	8669.353	-2.481	0.013116 *
isAnswerDarker	-71.206	5.040	8679.698	-14.128	<2e-16 ***
isAnswerLater	-23.591	7.118	8672.875	-3.314	0.000922 ***
isAnswerHigher	-14.866	7.106	8679.117	-2.092	0.036463 *
task:layout	26.359	14.129	77.358	1.866	0.065891
task:orientation	24.359	11.387	77.809	2.139	0.035551 *
layout:orientation	16.689	12.423	77.609	1.343	0.183045
task:layout:orientation	-139.786	12.301	77.512	-11.364	<2e-16 ***

Table 14: Output of the model extended with 5 more potential interference factors.

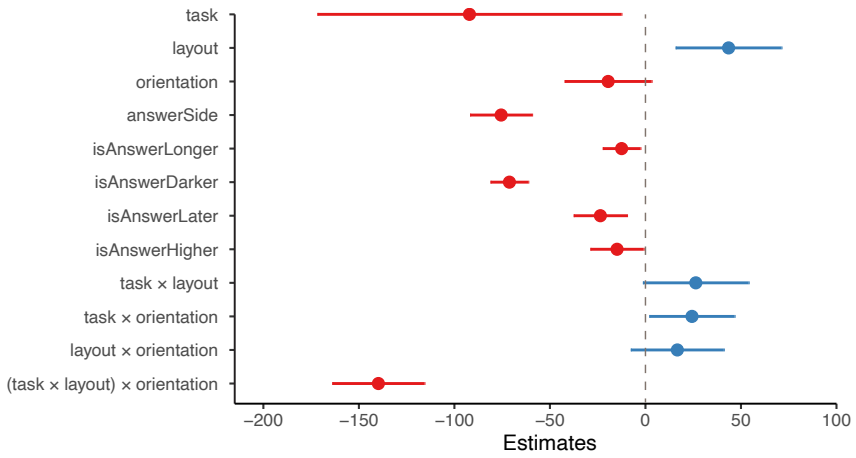

Figure 8: Coefficient estimate plot for model with fixed effects of all 5 potential interference factors. Error bars represent 95% confidence intervals. Negative coefficient indicates faster response time on tasks whose target answer bar is also the longer, darker, later, or higher bar. 3-way interaction is still significant after accounting for the effect of all the interference.

Extended model within height task:

$$\begin{aligned} \text{RT} \sim\ & \text{layout}{*}\text{orientation} + \text{answerSide} + \text{isAnswerLonger} \\ & + \text{isAnswerDarker} + \text{isAnswerLater} \\ & + (1 + \text{layout}{*}\text{orientation} + \text{answerSide} \mid \text{participant}) \end{aligned}$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1458.336	61.244	39.018	23.812	<2e-16 ***
layout	16.789	21.318	39.003	0.788	0.4357
orientation	-46.205	18.335	39.058	-2.520	0.0159 *
answerSide	-84.258	12.634	39.361	-6.669	5.86e-08 ***
isAnswerLonger	15.449	7.715	4327.861	2.002	0.0453 *
isAnswerDarker	-74.681	7.725	4329.556	-9.668	<2e-16 ***
isAnswerLater	-23.748	7.716	4328.709	-3.078	0.0021 **
layout:orientation	158.442	18.941	39.125	8.365	3.03e-10 ***

Table 15: Output of the extended model within height task. The factor "isAnswerHigher" is excluded as it is always true in the height task.

Extended model within index task:

$$\begin{aligned} \text{RT} \sim\ & \text{layout}{*}\text{orientation} + \text{answerSide} + \text{isAnswerLonger} \\ & + \text{isAnswerDarker} + \text{isAnswerHigher} \\ & + (1 + \text{layout}{*}\text{orientation} + \text{answerSide} \mid \text{participant}) \end{aligned}$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1254.782	53.404	39.020	23.496	<2e-16 ***
layout	70.696	18.561	38.186	3.809	0.000493 ***
orientation	7.450	14.154	38.983	0.526	0.601626
answerSide	-66.652	10.749	39.144	-6.201	2.69e-07 ***
isAnswerLonger	-40.162	6.425	4340.919	-6.251	4.47e-10 ***
isAnswerDarker	-68.064	6.451	4349.349	-10.551	<2e-16 ***
isAnswerHigher	-14.683	6.430	4347.344	-2.284	0.022438 *
layout:orientation	-124.739	16.244	38.368	-7.679	2.84e-09 ***

Table 16: Output of the extended model within index task. The factor "isAnswerLater" is excluded as it is always true in the index task.

Extended model within each pair of task and bar orientation:

$$\text{RT} \sim \text{layout} + \text{answerSide} + \text{isAnswerLonger} + \text{isAnswerDarker} + [\text{isAnswerHigher or isAnswerLater}] + (1 + \text{layout} \mid \text{participant})$$

Note that isAnswerHigher is excluded in the height task because the variable is always true in the height task. Similarly, isAnswerLater is excluded in the index task because the variable is always true in the index task.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1504.79	69.60	39.00	21.620	<2e-16 ***
layout	-141.35	31.43	39.06	-4.497	6.02e-05 ***
answerSide	-78.72	11.04	2200.12	-7.132	1.33e-12 ***
isAnswerLonger	5.75	11.00	2199.14	0.523	0.6012
isAnswerDarker	-58.92	11.01	2199.33	-5.353	9.57e-08 ***
isAnswerLater	-31.98	11.01	2199.46	-2.906	0.0037 **

Table 17: Output of extended model within the height task and horizontal bar conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1413.76	57.33	39.13	24.661	2e-16 ***
layout	177.16	24.93	39.05	7.107	1.52e-08 ***
answerSide	-90.96	11.07	2161.72	-8.214	3.64e-16 ***
isAnswerLonger	24.38	10.98	2158.87	2.221	0.0264 *
isAnswerDarker	-90.69	10.99	2159.23	-8.251	2.69e-16 ***
isAnswerLater	-14.77	10.97	2158.09	-1.346	0.1783

Table 18: Output of the extended model within height task and vertical bar conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1247.485	54.605	39.030	22.845	<2e-16 ***
layout	194.398	26.029	38.518	7.469	5.30e-09 ***
answerSide	-56.645	8.950	2179.828	-6.329	2.98e-10 ***
isAnswerLonger	-49.425	8.931	2179.720	-5.534	3.51e-08 ***
isAnswerDarker	-74.430	8.949	2180.026	-8.317	<2e-16 ***
isAnswerHigher	-14.047	8.944	2183.104	-1.571	0.116

Table 19: Output of the extended model within index task and horizontal bar conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1263.330	55.891	39.012	22.603	<2e-16 ***
layout	-54.933	23.216	36.264	-2.366	0.023432 *
answerSide	-78.344	9.393	2193.073	-8.341	<2e-16 ***
isAnswerLonger	-31.440	9.371	2192.497	-3.355	0.000807 ***
isAnswerDarker	-62.805	9.435	2197.541	-6.657	3.53e-11 ***
isAnswerHigher	-15.858	9.373	2192.803	-1.692	0.090812 .

Table 20: Output of the extended model within index task and vertical bar conditions.

Extended model within each pair of task and layout:

$$\text{RT} \sim \text{orientation} + \text{answerSide} + \text{isAnswerLonger} + \text{isAnswerDarker} + [\text{isAnswerHigher or isAnswerLater}] + (1 + \text{orientation} \mid \text{participant})$$

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1480.54	60.00	39.12	24.674	<2e-16 ***
orientation	116.31	22.84	39.20	5.094	9.21e-06 ***
answerSide	-125.80	11.27	2144.00	-11.163	<2e-16 ***
isAnswerLonger	13.19	11.16	2140.51	1.181	0.23756
isAnswerDarker	-68.72	11.18	2140.35	-6.149	9.30e-10 ***
isAnswerLater	-29.81	11.15	2141.12	-2.672	0.00759 **

Table 21: Output of the extended model within height task and stacked layout conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1440.43	69.16	39.00	20.827	<2e-16 ***
orientation	-202.74	28.98	38.96	-6.996	2.18e-08 ***
answerSide	-45.46	10.80	2218.73	-4.209	2.67e-05 ***
isAnswerLonger	17.45	10.77	2218.35	1.620	0.1053
isAnswerDarker	-81.98	10.78	2218.78	-7.605	4.17e-14 ***
isAnswerLater	-17.99	10.77	2218.79	-1.669	0.0952 .

Table 22: Output of the extended model within height task and side-by-side layout conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1324.932	59.922	39.017	22.111	<2e-16 ***
orientation	-116.918	23.203	37.934	-5.039	1.18e-05 ***
answerSide	-54.159	9.618	2176.765	-5.631	2.03e-08 ***
isAnswerLonger	-51.802	9.594	2175.654	-5.400	7.41e-08 ***
isAnswerDarker	-68.280	9.637	2179.459	-7.085	1.87e-12 ***
isAnswerHigher	-17.580	9.590	2175.375	-1.833	0.0669 .

Table 23: Output of the extended model within index task and stacked layout conditions.

	Estimate	SE	df	t-value	Pr(> |t|)
(Intercept)	1184.817	52.891	39.002	22.401	<2e-16 ***
orientation	133.050	19.748	37.566	6.737	5.91e-08 ***
answerSide	-79.807	8.716	2197.362	-9.156	<2e-16 ***
isAnswerLonger	-29.431	8.701	2196.916	-3.383	0.00073 ***
isAnswerDarker	-68.654	8.735	2200.039	-7.860	5.96e-15 ***
isAnswerHigher	-12.533	8.718	2202.046	-1.438	0.15069

Table 24: Output of the extended model within index task and side-by-side layout conditions.

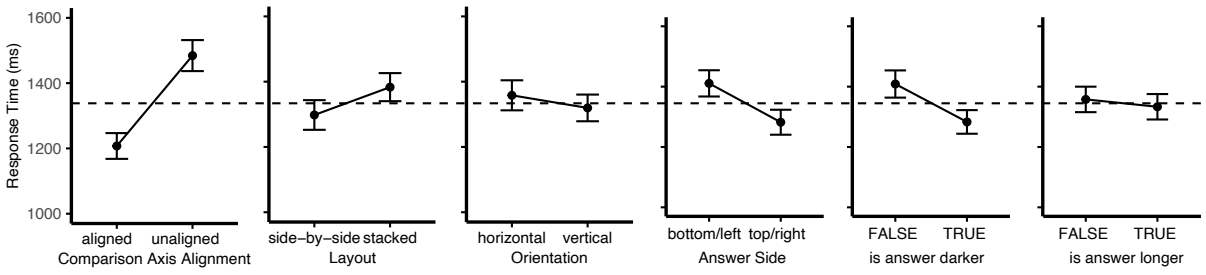

Figure 9: Mean response times grouped by each factors including the potential interference factors. Response time was faster when top or right chart was the answer. Response time was faster when the target answer bar was also the darker bar.

4 Exploratory Analysis

4.1 Accuracy

Note that higher means better in figures with accuracy with measure, unlike those with reponse time as measure.

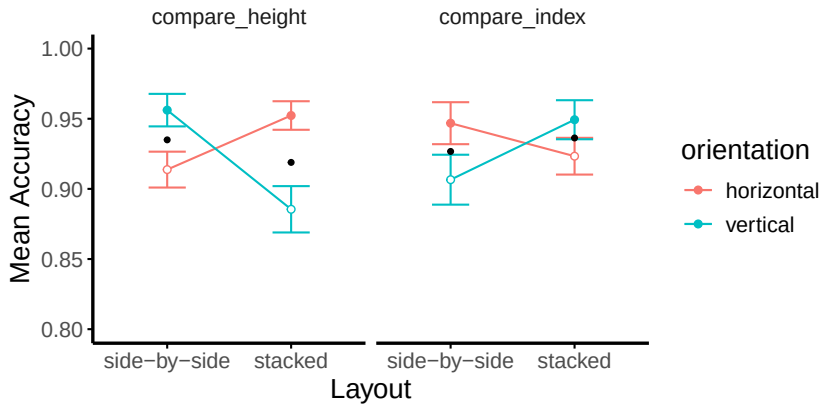

Figure 10: Mean accuracy of participants by task, layout, and orientation. Filled dots represent conditions where comparison axis is aligned. Black dots represent the mean of each task-layout pair. Error bars represent the standard error of the mean. We can see the similar interaction pattern with that of response time analysis. Aligned comparison axis yields the highest accuracy.

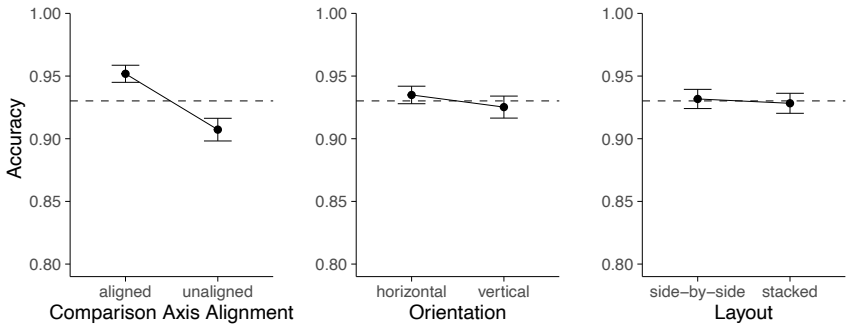

Figure 11: Effects of comparison axis alignment, layout, and orientation on participants' mean accuracy. Dashed line represents mean of each participant's mean accuracy. Error bars represent the standard error of the mean. Comparison axis alignment has the strongest effect on accuracy.