

WCST data analysis - Step 2

Manipulate data

Read data frames and manually rated data

We have loaded data frames saved by the Step 1 script, and merged the manually categorized free text answers about the rule with participantdata_inc.

Exclude participants

We excluded extra participants if there were more than 78 in a condition.

Modify data

Difficulty of tasks was supposed to be a number 1-10, but it was also a free text answer, and some participants went outside the recommended range (e.g.: difficulty = 50). We capped these numbers to 10.

Descriptive statistics

The number of participants was 78 in each condition.

Condition	Number of participants	Number of solvers	Solution rate	Median task time (min)	Aha rate of so
Letters In	78	78	1.00	1.16	
Letters Out	78	73	0.94	2.12	
No Letters	78	50	0.64	8.24	
Fixed Sequence	78	69	0.88	4.80	
Ambiguous Cards	78	40	0.51	13.75	
Moon Card	78	70	0.90	3.20	
Letters Only	78	78	1.00	0.70	

Statistical tests

Difficulty of the task

A task is more difficult in one condition than in another condition, if less participants are able to solve it, or if it takes longer for participants to solve it (even if solution rates are the same). This is why, we compared the difficulty of the task

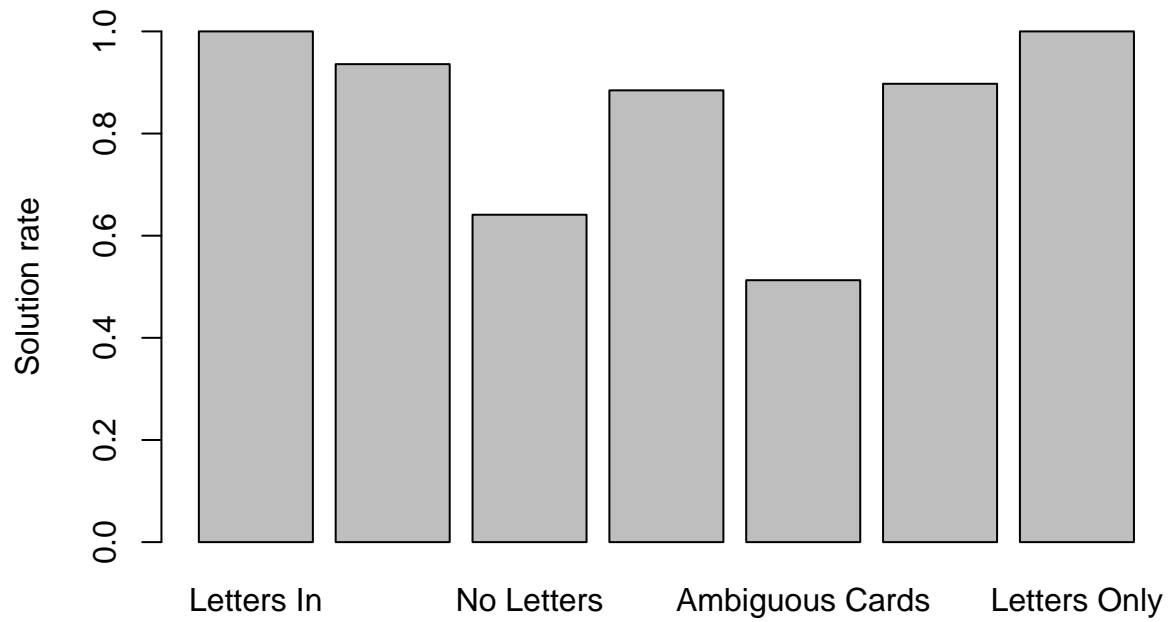
- first, by Fisher's exact test on the number of solvers (pairwise comparisons)
- then, if the Fisher's exact test was not significant, we compared task time by two-sample Wilcoxon test (same as the Mann-Whitney test) or by
- ANOVA with Greenhouse-Geisser correction: robust also for non-normally distributed data

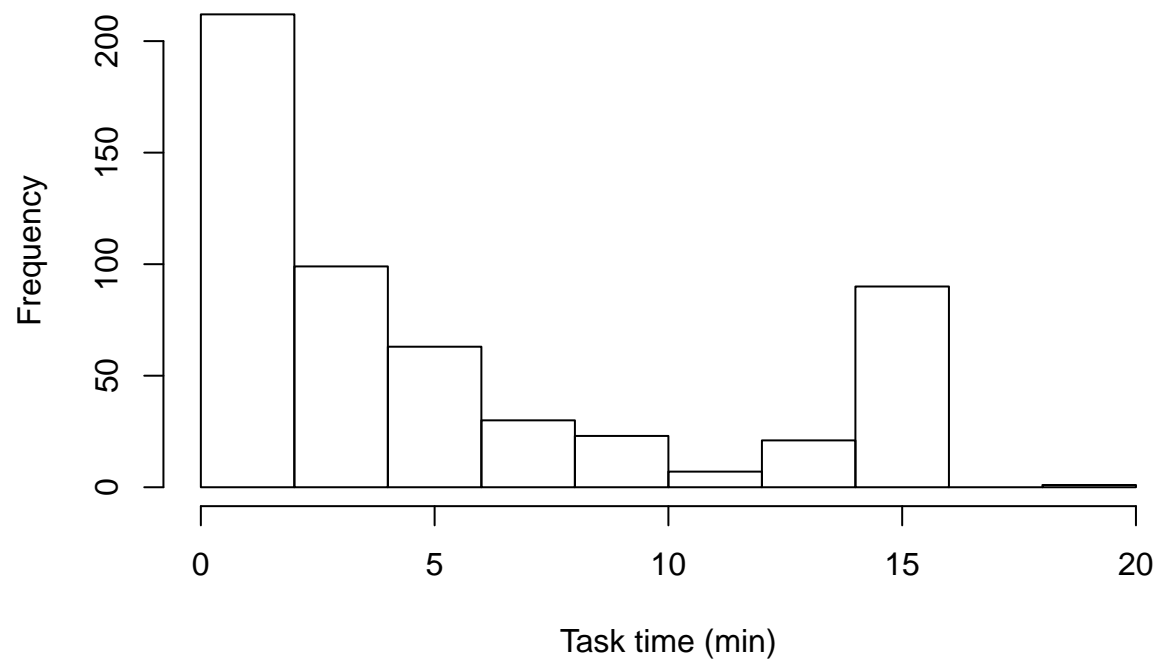
We also used binary logistic regression.

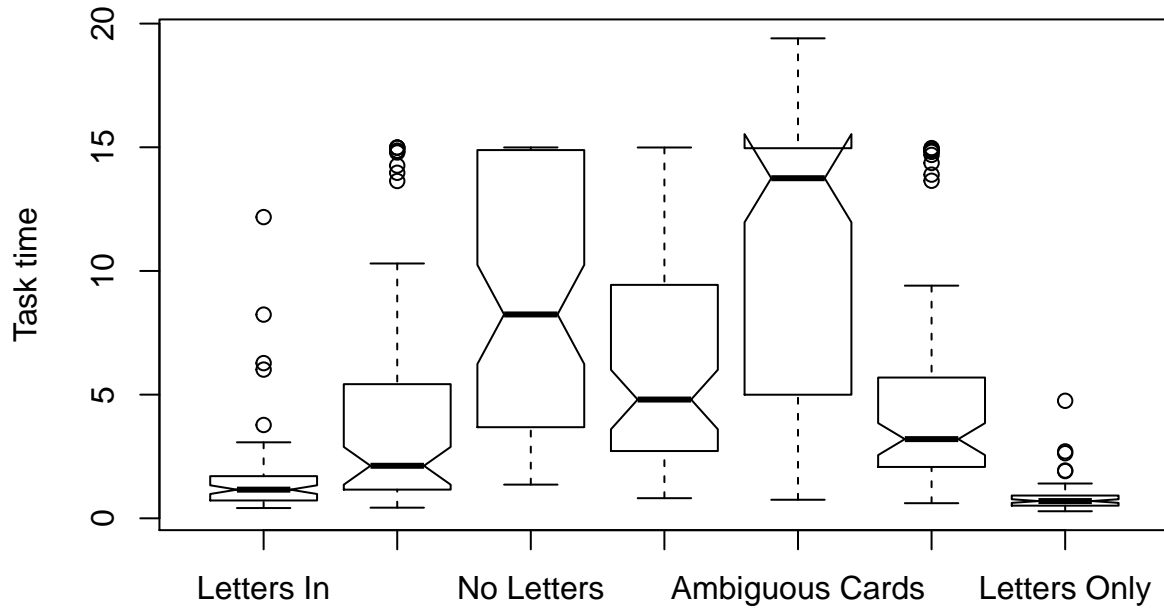
The following plots compare all conditions of Experiment 1 and 2 and the control condition:

- The barplot shows solution rates accross conditions.

- The histogram shows the distribution of task time in all the conditions.
- The boxplot shows the distribution of task time accrosss condition. Circles represent outliers. Whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. If the notches of two boxes do not overlap this is strong evidence that the two medians differ.







The histogram of task times shows that the data is not normally distributed, so we had to use nonparametric tests. The Kolmogorov-Smirnov normality test on task time also showed that the data was not normally distributed.

```
c(D = 0.655758918857444), 0, two-sided, One-sample Kolmogorov-Smirnov test, participantdata_inc %>%
select(task_time)
```

Aha feelings

We used Fisher's exact tests to make pairwise comparison between conditions regarding the Aha-feelings of solvers (whether they reported having aha-feelings or not).

Results - Experiment 1

Difficulty of the task - Experiment 1

We analyzed the contingency table containing the number of solvers and non-solvers in pairs of conditions with Fisher's exact test. A $p < 0.05$ means that the row/column association is statistically significant:

WLIN-WLOUT: $p = 0.0585073$

WLOUT-WNOL: $p = 8.2057452 \times 10^{-6}$

Wilcoxon test:

WLIN-WLOUT: $p = 5.0546397 \times 10^{-7}$

Aha feelings - Experiment 1

We analyzed the contingency table containing the number of solvers who reported Aha-feelings and number of solvers who did not report Aha-feelings in pairs of conditions with Fisher's exact tests. A $p < 0.05$ means that the row/column association is statistically significant:

WLIN > WONLY: $p = 0.001533$

WLOUT > WONLY: $p = 4.4032686 \times 10^{-7}$

WNOL > WONLY: $p = 7.2365593 \times 10^{-8}$

WLOUT > WLIN: $p = 0.0500146$

WNOL > WLOUT: $p = 0.3993585$

Results - Experiment 2

Difficulty of the task - Experiment 2

MOONSQL < WNOLA: Removing distracting visual cues makes finding the sequence rule easier
WNOLFS < WNOLA: The sequence rule can be used in both conditions but the exclusion rule can only be used in the WNOLFS condition, which might help.

We analyzed the contingency table containing the number of solvers and non-solvers in pairs of conditions with Fisher's exact test. A $p < 0.05$ means that the row/column association is statistically significant:

WNOLFS-WNOLA: $p = 5.5523537 \times 10^{-7}$

MOONSQL-WNOLA: $p = 1.6411398 \times 10^{-7}$

Aha feelings - Experiment 2

WNOLFS > WONLY: $p = 3.7576042 \times 10^{-5}$

WNOLA > WONLY: $p = 2.9906516 \times 10^{-4}$

MOONSQL > WONLY: $p = 0.0241923$

Other ideas

Language

Restructuring

Sliding average for move time

Rules used (matching rule for moves)

Correct rule

Does it correlate with problem difficulty or aha

Aha-rate of non solvers

Correlates with difficulty? Number of false ahas vs subjective difficulty or solution time

Goal of the experiment