WCST analysis

Data

Data sources

To run this script, copy-paste the following files to "datadir", and change "datadir" to your path to that folder:

- prolific_export_5f917030b1ac5a05a2123cac.csv: this contains the demographic data of participants from Prolific
- all the result files from the server (one for each participant): these contain the data from the game (pattern: condition_ID_Results.txt)
- vpn.txt from the server: this contains data from the pre-game questionnaire
- survey.txt from the server: this contains data from the post-game questionnaire
- time.txt from the server: this contains the timestamps of different action from the participants during the experiment (e.g., started game, filled out survey, etc.). This is not used for data analysis, but can be useful to check manually if something is not clear

We have manually made some modifications to these files to ensure that the data is read correctly. In the vpn.txt file:

- Participant 306: accidentally copied an URL next to their ID, we deleted the URL
- Participant 509: wrote "4o" for age, we corrected it to 40
- Participant 894: copy-pasted a random text to the ID field, we deleted it

In the survey.txt file:

- We deleted empty lines, where participants hit enter in a free text field for the following participants: 188, 237, 298, 363, 434, 529, 563, 582, 646, 763, 790
- Participant 593 wrote "10!" for difficulty, we corrected it to 10
- Participant 790, when describing the rule the html number > appeared insead of ">"; we changed it back, so the rule reads "1>2>3>4."
- Participant 886 wrote "d" for the level of difficulty field - we deleted this entry
- We collected survey data separately for a few participants for whom the server froze during their survey. We made a google form with the survey questions and we asked participants to fill it out after they indicated their problem via the Prolific messaging system. This usually happened within a week of them completing the experiment. We excluded those answers where participants indicated that they do not remember the experiment well. Some participants proactively sent us their survey answers via messaging within the Prolific platform before we could send them the link for the google form. For these participants, we copy-pasted their answers from their message to the Google form.

We merged the data from these files into a data frame, called "participantdata".

Variables

For each move participants made during the game we had the following variables:

- Trial number
- Source card
- Target card
- Match: correct or incorrect move

- Total time: the time elapsed between the source card appearing and the source card being dropped
- Move time: the time elapsed between dragging and dropping the source card
- Think time: total time - move time
- Rule: which rule the move conformed to (color, number, shape, index)

For each participant we calculated the following variables:

- Number of moves
- Task time: Time the participant spent with the task (min)
- Experiment time: The the participant spent with the entire experiment, including the surveys (min)
- Mean total time: mean time between moves (ms)
- Mean move time: mean time between drags and drops (ms)
- Mean think time: mean time the participant spent thinking between moves (total time - move time; ms)

Exclusion criteria

We excluded participants who:

- Quit during the first data sheet: 148
- Indicated that they were colorblind: 5
- Did not start the game after filling out the data sheet: 81
- Refreshed the screen during the game or went back to the instructions page after starting the game: 33
- Revoked their consent to use their data: 53
- Played the game more than once, even if they got assigned a different condition after the first time (we still included their first attempt though): 11
- Who did not fill out the final survey: 54

Those participants did not fill out the survey could be those, who

- quit the experiment during the game before time out, or
- left their computer and never pressed the "Continue" button after time out, or
- who finished their game, but who could not submit the survey, because the server froze.

These participants did not complete the survey, but still have some data from the game.

For several participants, the screen froze on the survey page. For these participants we sent out the survey questions separately in a Google form. 40 participants completed the form, so we could include their data in the analyses.

Our participant funnel looked like this:

- The number of experimental sessions started: 940
- The number of participants who gave informed consent and filled out the first datasheet: 792
- The number of participants who indicated that they were not colorblind: 787
- The number of participants who started the game: 706
- The number of participants who did not refresh the game or went back to the instructions page: 673
- The number of participants who did not revoke their consent to use their data: 620
- The number of participants who only played the game once: 609
- The number of participants who did not give up during the game and filled out the last survey too: 555

All-in-all, we excluded 385 initialized experimental sessions and we were left with 555 participants.

We need 78*7 = 546 participants. 546 - 555 = -9 participants needed.

To do on Prolific

Approve the participation and pay the following participants (to be copied to Prolific -> Approve by upload):

Batch approve list

Manually reject participation on Prolific:

- If the same ID shows "quit" and then "played twice", approve anyway. These are probably cases where the game froze and then the participant restarted the experiment.
- Check Messages too, because sometimes participants explain their technical difficulties there.
- Check comments in the survey too, because sometimes it is obvious from their comments that they finished the experiment, even though they have missing data.

prolific_ID	prolific_action

Costs

We estimated that the median time for completing the experiment would be 8 minutes. The suggested hourly rate is 7.5 GBP on the Prolific website. From this, we calculated the fee for the task to be 1 GBP (independently of how long it took for the participant to finish). Together with 33% service fee and 20% VAT of the service fee, each approved participant costs around 1.396 GBP.

The time-out limit was calculated by the Prolific website to be 39 minutes for the whole experiment including the surveys (for the game, our own time-out was 15 minutes).

819 participants initiated the experiment on the Prolific website. From this:

- 110 participants returned the task
- 17 participants timed-out
- We excluded 31 participants for various reasons (see above)
- We approved and paid the remaining 661 participants
- From these, we could use the data of 555 participants

Sometimes we lost data because of server error. Other times, the browser of participants froze (probably also server error). In these cases, we cannot use the data of these participants, but we still have to pay them, because it was not their fault.

Costs for the experiment should be: 922.756

The pilot experiments cost 73.23 GBP all together:

- Pilot 1: 33.60 GBP
- Pilot 2: 16.80 GBP
- Pilot 3: 14.01 GBP
- Pilot 4: 8.82 GBP

Analyses

Stopping criteria and further data cleaning:

We run the experiment until we had enough (78) participants in each condition after automatically evaluating them based on the exclusion criteria described above. First, the conditions were assigned randomly, then as we got closer to the desired number of participnats, we only let a certain number of participants to enter each condition to make up for the missing number of participants. This process was not exact, so we ended up with a few extra participants in some conditions. We stopped the experiment when each condition had at least 78 participants after the exclusions.
We read the survey answers of the remaining participants and decided to exclude a few more:

- . . .

After this manual excludion process,should we exclude the extra participants if there are more than 78 in a group?

```
## Warning in `!=.default`(prolific_ID, exclude_list): longer object length is not
## a multiple of shorter object length

## Warning in is.na(e1) | is.na(e2): longer object length is not a multiple of
## shorter object length
```

Summary of conditions

Condition	Number of participants	Aha rate of solvers	Failure rate	Avg task time	Avg number of moves
wlin	85	0.800	0.000	1.584	31.082
wlout	78	0.932	0.064	4.345	70.103
wnol	79	0.980	0.354	8.907	165.190
wnolfs	78	0.884	0.115	6.606	124.962
wnola	78	0.900	0.487	10.357	234.577
moonsq	79	0.761	0.101	4.659	95.620
wonly	78	0.577	0.000	0.868	19.308

We need 78 participants per condition.
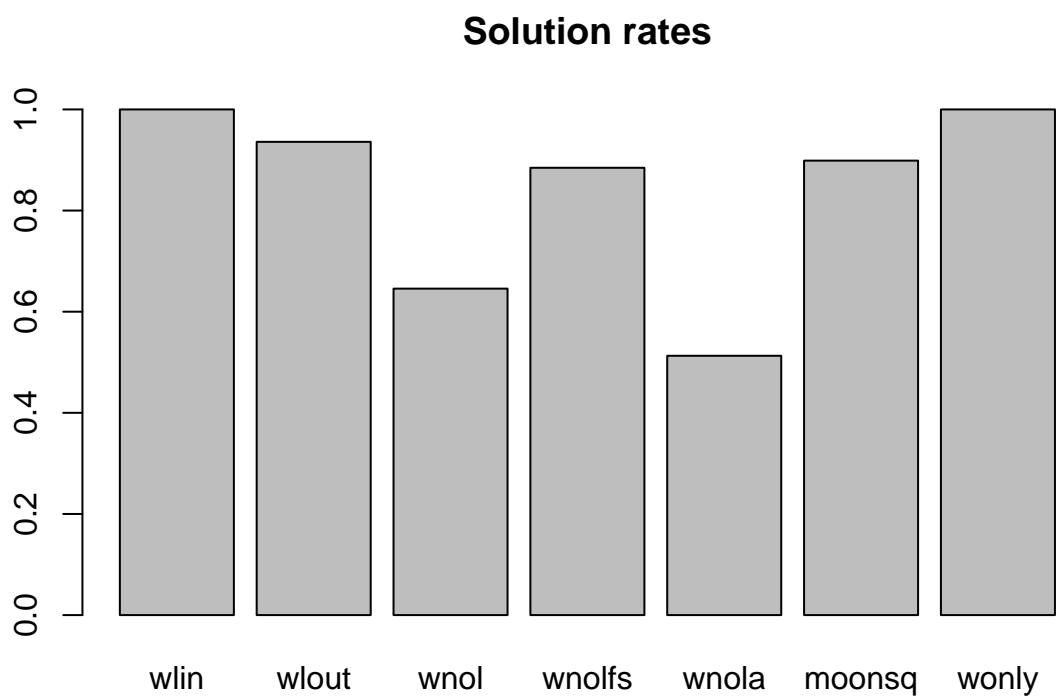**We need 0 more participants**
That is 0 GBP
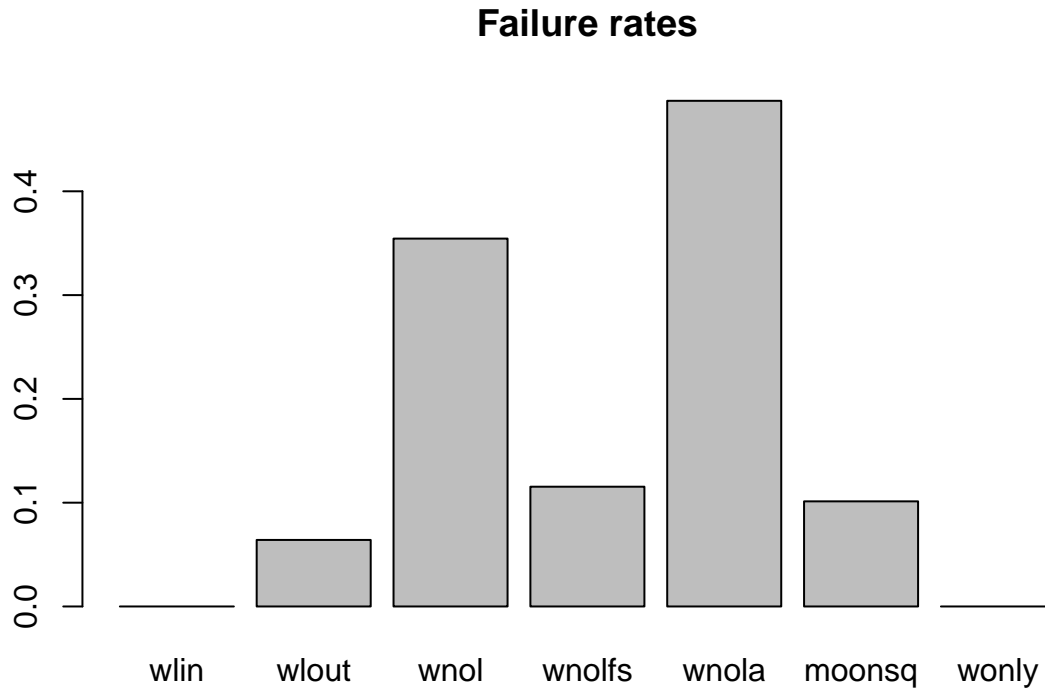
Participant plots

Difficulty of the task

Solution rate: Fisher's exact tests

We analyzed the contingency table containing the number of solvers and non-solvers in pairs of conditions. A p<0.05 means that the row/column association is statistically significant.
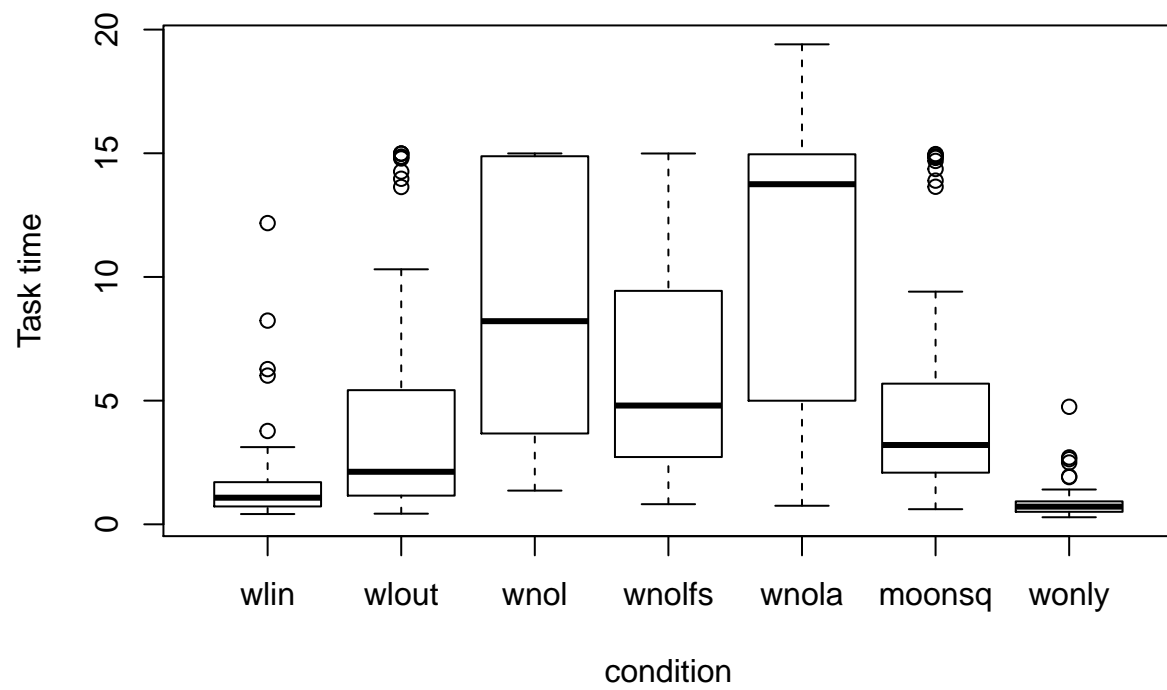
Solution rates

Failure rates

Solution time: ANOVA

We checked whether the data was normally distributed with Kolmogorov-Smirnoff test:

If the data is normally distributed, we use ANOVA, if it is not, we use Wilcoxon.

Aha feelings: Fisher's exact test

Aha rate of solvers