Data Processing Exp3

#01 projects/exp3# #active/priority

Clean up steps for this

1. Create generic RMD and Q files
2. place all needed R and Python files in one directory

Step 1 - Preprocessing data files from Presentation and output to csv file

Presentation outputs .log files and .txt files. You will need to Preprocess both in Python before cleaning the data in R.

1. Copy participant data file to directory with “Raw_to_excel/ExpX_raw” folder - this is for archive
2. Copy the 18 participant data .txt and .log files to “P_work” folder in the same directory
3. Open these 2 python files in the “Python_files_IAA_Cleaning” folder
    1. Exp_3_time/Python_files_IAA_Cleaning/IAA_log_splitting_rename.py
    2. Exp_3_time/Python_files_IAA_Cleaning/IAA_txt_splitting_rename.py
4. Directory should point to P_work folder for list and source
5. Run each file separately - these files will populate the "Exp3_csv/P_txt " and the “Exp3_csv/P_log” folders, ultimately placing 14 csv files in “Exp3_csv/P_complete folder.
6. Confirm the folder P_complete folder contains 2 sets of B1 - B6 + BT files - one with Pz_ prefix and one set with Txt_prefix
7. create new folder using participant number with _csv suffix.
8. Move to the Exp3_csv folder for archive

Step 2 - Clean the csv using R scripts

8. Copy new participant number folder to the Exp3 cleaning folder and place 3 R scripts inside the folder
    1. Exp3_csvCombo_1_PVAR.R - creates cleaned csv
    2. P163XX_X_Descriptives_Dual_Exp3.Rmd - creates summary of participant
    3. P163XX_Eligibility.qmd
9. Set R studio working directory to current P_folder and open the three files listed in step 8
10. Run the files in this order
    1. combo file ➔ Input new pariticpant number and version number in lines 68 & 69 then run all
        - note file stops midway through for processing duplicate responses - follow in file directions (copied below) - there is a deliberate break CNTRL +F (open find) type “ halt “ to get to the right place ln1004
        1. in global environment check “dupesX_are” for observations
        2. If there are no observations for duplicates go to Step XX for evey block that has duplicates you need to follow these steps

1. lines 1007 - 1015 ➜ remove hash tags from 1007 & 1015
2. remove hash tags from the dupes_are with observations
3. highlight lines 1007 - 1018 and run ➜ creates dupes all csv and df
4. open the dupes_all df ➜ open each df listed in turn and do the following
    1. open the corresponding “detect_X” df
    2. scroll down to the trial number listed in “dupes_all”
    3. note the duplicate response row numbers to be removed according to the following criteion
        1. response post target onset - keep first reponse, remove any additions
        2. if there is a response before trial type AND is occurs < 600ms from onset remove that response and keep response post target onset
        3. if there is a response after 600ms onset keep this response and remove response post target onset
    4. lines 1022 - 1028 - contain preset code to remove duplicated response lines by block. Insert the appropriate line to remove and run the line
    5. reopen df and cross check the trial is now as required
    6. repeat for each duplicate response
3. Run the rest of the file (thinking about splitting these into 3 - up to the duplicates - then up to combo csv creation then the final assessment

2. run the rmd ➜ type participant number in line 2 & hit knit
3. run the qmd ➜ type participant number below title (bolded) and in last chunlk
    1. then type the participant number + ver to read teh Rdata file
    2. hit render

11. look at the qmd file for performance exclusions ➜ > 70% ACC during exposure
12. Take the combo excel file and place it in the directory with the other participant combo files
13. when you have the participants assmebled, then run the xxx file to create your data set
14. use the R script to merge demographics and duplicates csvs

Step 3 create the combined df

1. Set R directory to ~/Documents/R Stuff Cupboard/Exp_3_time/Combo_Section
2. open Ex3_VAR_Combine_Clean_Indys.R
    1. line 69 input subject number range
    2. line 70 input versions
    3. line 71 input number of participants as a number
3. ensure ALL of participants you wish to include have a “P163XX_X_COMBO_Cleaned.csv” csv file in the directory
4. ensure there are no other csv files in the directory

5. run the file
6. reset directory to same folder and open Exp3_Combo_Prelim_Eligibility.qmd and P163xx_x_Descriptives_Dual_Exp3.Rmd
7. Update P range line 2 of the .rmd file and hit knit
8.