

Persuading to Prepare for Quitting Smoking with a Virtual Coach: Using States and User Characteristics to Predict Behavior

Overview of Steps to Reproduce Analyses

Author: Nele Albers

Date: February 2023

This file is meant to give you an overview of the steps needed to reproduce our analyses from the paper "Persuading to Prepare for Quitting Smoking with a Virtual Coach: Using States and User Characteristics to Predict Behavior" by Nele Albers, Mark A. Neerincx, and Willem-Paul Brinkman. All code was written and run by the first author Nele Albers.

Types of Analyses

Below we describe how you can reproduce our results.

Data and Preprocessing

The folder "Data" contains the data we used for our analysis:

- The file "_Explanation_of_Data_Files.xlsx" contains a description of each data file.
- There is a separate data file for each study component. This includes data we extracted from participants' Prolific profiles (prescreening_prolific_numeric_anonym.csv), data from a prescreening questionnaire (prescreening_questionnaire_anonym.csv), and data from a pre-questionnaire, the conversational sessions, and the post-questionnaire.
- For each data file, there is an explanation file that explains each column of data. For example, the file "post_questionnaire_data_explanation.xlsx" explains the data columns from the data file "post_questionnaire_anonym.csv."
- All data has been anonymized. There is a new random identifier "rand_id" that can be used to link the separate data files.

For the sake of completeness, we also include data on people's assignment to the four study groups in the longitudinal study that was used to collect the data ("Data/assignment_anonym.csv"). The four study groups determined how people were persuaded after the second session in the study. The policies for the four study groups are described in the Appendix of the paper.

The file "extract_RL_samples.py" is used to further preprocess the gathered data. Running the file will:

- re-create the output-file "extract_RL_samples_output.txt" that contains the Cronbach's alpha values for index measures as well as the minimum and maximum age of participants.
- re-create the two data files in the folder Analysis/Data that are used for the analyses.

We have already run this preprocessing code, so there is no need for you to run it to be able to run the analysis code.

Analysis for the 6 Research Questions and Participant Characteristics

Navigate to the folder "Analysis" to reproduce our results for the six research questions as well as our reported participant characteristics from the Appendix. The "Analysis"-folder contains a README-file with further instructions.

Explanation of Files and Folders

This folder contains the following folders and files:

- Analysis: Code and instructions for reproducing our results for the six research questions as well as the participant characteristics reported in the Appendix.
- Data: Data from the separate study components.
- extract_RL_samples.py: Code to further preprocess the data from the study components for our analysis.
- extract_RL_samples_output.txt: File that shows the output generated during the preprocessing. This includes the Cronbach's alpha values for index measures and the minimum and maximum age of participants.
- README.md/README.pdf: This README-file.