

Data underlying the analysis for study 2

Author: Nele Albers

Date: June 2024

This README-file is meant to give you an overview of the data underlying the analysis for study 2.

The data provided here is a subset of the data provided in this repository:

<https://doi.org/10.4121/1d9aa8eb-9e63-4bf5-98a3-f359dbc932a4>.

Types of information

We provide the types of information explained below.

Raw data

Navigate to the folder "Raw_Data" to inspect the anonymized raw data from the study used in our analysis. This includes:

- Data from participants' Prolific profiles (e.g., age, gender):
prolific_profile_multistudypaper_anonym.csv,
- Data from the prescreening questionnaire (e.g., smoking frequency):
prescreening_questionnaire_multistudypaper_anonym.csv, and
- Data from the conversational sessions with the virtual coach (e.g., effort spent on activities): sessionsdata_multistudypaper_anonym.csv.

For each data file, there is a corresponding .xlsx-file that explains each data column. For instance, the file

"prolific_profile_multistudypaper_data_explanation.xlsx" explains the data contained in the file "prolific_profile_multistudypaper_anonym.csv."

We further provide the code used to create these anonymized files. For example, the file "anonymize_sessiondata_multistudypaper.py" shows how we anonymized the data from the conversational sessions. Since we cannot provide the non-anonymized data, you cannot run these files yourself.

Merged data

In our analysis, we used the merged data files

"effort_data_all_activities_human_inv_multistudypaper.csv" and

"merged_data_human_inv_multistudypaper.xlsx." The former contains the effort ratings for all activities; the latter contains the effort ratings and activity experiences for all future self-related activities together with corresponding participant characteristics. The file "preprocess_sessiondata_multistudypaper.py" shows how we created these two files from the raw data files.

Mean effort per activity

We also computed the mean effort per activity using the file

"compute_mean_effort_per_activity.py." The results are shown in the file

"compute_mean_effort_per_activity_output.txt." To see which activities the indices correspond to, refer to this file: https://github.com/PerfectFit-project/virtual_coach_human_inv/blob/main/actions/Activities.xlsx

published as part of the implementation of the virtual coach Kai used in the study.

Explanation of files and folders

This directory contains the following files and folders:

- Raw_Data:
 - anonymize_sessionsdata_multistudypaper.py: Code used to anonymize the data from the conversational sessions.
 - clean_and_anonymize_prescreening_data_multistudypaper.py: Code used to anonymize the data from the Prolific profiles and prescreening questionnaire.
 - clean_and_anonymize_prescreening_data_multistudypaper_output.txt: Log-file created when running the file above with some information on data removed during cleaning.
 - prescreening_questionnaire_multistudypaper_anonym.csv: Anonymized data from the prescreening questionnaire (e.g., smoking frequency).
 - prescreening_questionnaire_multistudypaper_data_explanation.xlsx: Explanation of the data columns in the file above.
 - prolific_profile_multistudypaper_anonym.csv: Anonymized Prolific profile data of participants (e.g., age).
 - prolific_profile_multistudypaper_data_explanation.xlsx: Explanation of the data columns in the file above.
 - sessionsdata_multistudypaper_anonym.csv: Anonymized data from the conversational sessions with the virtual coach (e.g., effort spent on activities).
 - sessionsdata_multistudypaper_data_explanation.xlsx: Explanation of the data columns in the file above.
- compute_man_effort_per_activity.py: File to compute the mean effort per activity.
- compute_mean_effort_per_activity_output.txt: File containing the mean effort per activity and across all activities together as the result of running the file above.
- effort_data_all_activities_human_inv_multistudypaper.csv: File containing the effort for all activities.
- merged_data_human_inv_multistudypaper.xlsx: File containing the effort ratings and activity experiences for all 8 future self-related activities together with corresponding participant characteristics.
- preprocess_sessiondata_multistudypaper.py: File used to create the data files "effort_data_all_activities_human_inv_multistudypaper.csv" and "merged_data_human_inv_multistudypaper.xlsx."
- README.md/README.pdf: This README-file.