

# Data

**Author:** Nele Albers

**Date:** January 2025

This README-file is meant to give you an overview of the data underlying our analyses.

Authored by Nele Albers, Francisco S. Melo, Mark A. Neerincx, Olya Kudina, and Willem-Paul Brinkman.

## Types of information

Below we describe the types of information we provide in this repository.

### Activity difficulty data

The folder "Activity\_Difficulty\_Data" provides the preprocessed activity difficulty ratings data. To reproduce the preprocessing, first download the file "ratings\_data\_anonym.csv" from this DOI: <https://doi.org/10.4121/5198f299-9c7a-40f8-8206-c18df93ee2a0>. Place the downloaded file in the "Activity\_Difficulty\_Data"-folder.

### Data from previous studies without human feedback

The folder "Data\_Previous\_Studies" should store the effort scores from our two previous studies without the possibility of receiving human feedback:

- The data from the first study needs to be downloaded from this DOI: <https://doi.org/10.4121/21533055>. Specifically, you need to download the file "conversational\_sessions\_anonym.xlsx" and save it as "data\_study\_0\_conversational\_sessions\_anonym.xlsx."
- The data from the second study needs to be downloaded from this DOI: <https://doi.org/10.4121/9c4d9c35-3330-4536-ab8d-d5bb237c277d>. Specifically, you need to download the file "Step\_5\_Training\_The\_Model\_Results/Data/sessionsdata\_anonym.csv" and save it as "data\_study\_1\_sessionsdata\_anonym.csv."

### Anonymization code

We provide the code used to anonymize the data from our study. For example, the file "anonymize\_sessiondata\_human\_inv.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. For some of the preprocessing files, there is also a corresponding .txt-file with some log data. For example, the file "clean\_and\_anonymize\_postquestionnaire\_data\_output.txt" provides log information for running the file "clean\_and\_anonymize\_postquestionnaire\_data.py."

### Anonymized data from our study

We provide the anonymized data from our study. This includes:

- Data from the prescreening questionnaire (e.g., stage of change for quitting smoking),
- Data from people's Prolific profiles (e.g., age, gender),
- Data from the conversational sessions (e.g., effort spent on activities),
- Data from the post-questionnaire (e.g., weights for allocation principles), and
- Data on people clicking on the reading confirmation links in the human feedback messages.

For each data file, there is a corresponding .xlsx-file explaining each column in the data file. For example, the file "sessionsdata\_data\_explanation.xlsx" explains the columns in the file "sessionsdata\_anonym.csv."

Data from the different files can be linked using the anonymized participant identifier "rand\_id."

## **Explanation of files and folders**

This directory contains the following files and folders:

- Activity\_Difficulty\_Data:
  - Activity\_Formulations\_Competencies\_Study\_and\_Mapping\_to\_Human\_Inv\_Activities.xlsx: Mapping between activity indices in our study and the activity difficulty ratings dataset.
  - merge\_data\_difficulty\_data\_for\_human\_inv\_study.py: Get the activity difficulty ratings for the activities used in our study.
  - Merged\_Data\_Difficulty\_Ratings\_Activities.csv: Activity difficulty ratings for the activities used in our study.
- Data\_Previous\_Studies:
  - This folder is initially empty. To reproduce our analyses, you need to download the two files with data from previous studies and save them in this folder as described earlier in this README-file.
- anonymize\_sessiondata\_human\_inv.py: Anonymize the data from the conversational sessions.
- clean\_and\_anonymize\_feedback\_reading\_confirmation\_data.py: Anonymize the data on people clicking on reading confirmation links in the feedback messages.
- clean\_and\_anonymize\_feedback\_reading\_data\_output.txt: Output of running the file above.
- clean\_and\_anonymize\_postquestionnaire\_data.py: Anonymize the post-questionnaire data.
- clean\_and\_anonymize\_postquestionnaire\_data\_output.txt: Output of running the file above.
- clean\_and\_anonymize\_prescreening\_data\_human\_inv.py: Anonymize the prescreening data (prescreening questionnaire + Prolific profiles).
- clean\_and\_anonymize\_prescreening\_data\_human\_inv\_output.txt: Output of running the file above.
- feedback\_reading\_confirmation\_anonym.csv: Anonymized reading confirmation data.
- feedback\_reading\_confirmation\_data\_explanation.xlsx: Explanation of the columns in the file above.
- postquestionnaire\_anonym.csv: Anonymized post-questionnaire data.
- postquestionnaire\_data\_explanation.xlsx: Explanation of the columns in the file above.
- prescreening\_questionnaire\_anonym.csv: Anonymized prescreening questionnaire data.
- prescreening\_questionnaire\_data\_explanation.xlsx: Explanation of the columns in the file above.

- prolific\_profile\_anonym.csv: Anonymized Prolific profile data.
- prolific\_profile\_data\_explanation.xlsx: Explanation of the columns in the file above.
- README.md/README.pdf: This Readme-file.
- sessionsdata\_anonym.csv: Anonymized data from the conversational sessions.
- sessionsdata\_data\_explanation.xlsx: Explanation of the columns in the file above.