TABLE OF CONTENTS

[Raw Data Structure 1](#_Toc115181000)

[Block1: 1](#_Toc115181001)

[Block 2 and 3: 2](#_Toc115181002)

[Variables of interest 2](#_Toc115181003)

[Block1: 2](#_Toc115181004)

[Block 2 and 3: 2](#_Toc115181005)

[Catch Trials/attention checks 2](#_Toc115181006)

[Data processing 3](#_Toc115181007)

[Cleaning and preparing the data 3](#_Toc115181008)

[Exp1 3](#_Toc115181009)

[Exp 2 3](#_Toc115181010)

[Running the Matlab Script 3](#_Toc115181011)

[Variable descriptions 4](#_Toc115181012)

[Data Trials 4](#_Toc115181013)

[Subject Level Data 4](#_Toc115181014)

[Data Analysis 5](#_Toc115181015)

# Raw Data Structure

The raw data are divided into three folders for each Exp.

*GitHub\Data and Scripts\Exp2\RawData\Data1*

*GitHub\Data and Scripts\Exp2\RawData\Data2*

*GitHub\Data and Scripts\Exp2\RawData\Data3*

Since there were three blocks in each experiment, the raw data files are divided accordingly in three files, within a Data file.

*Block1\_AccFirst1.xlsx*

*AccFirst1.xlsx*

*ShSecond1.xlsx*

## 

## Block1:

In block 1, participants observed 30 health-related statements in random order, each for 6 seconds. Statements were randomly selected for each participant from a list of 60 (see supplement material for all statements).

## Block 2 and 3:

In either the second or third block (counterbalanced across participants), participants observed 60 statements (half were new, half repeated) one at a time, in a random order. For each statement, they indicated whether they wanted to share the statement with participants who might play a similar task the following day.

In either the second block or third block (counterbalanced across participants), participants indicated whether they thought each statement was true or false.

## Variables of interest

Out of the raw data we obtained, the following variables were of use to answer our main research questions. Other variables as can be found in the raw data files were not used further.

### Block1:

**-**Trial number

-sentence

-sentence number

### Block 2 and 3:

**-**Trial number

-sentence

-sentence number

-accuracy or sharing rating

**Attention Checks:**

-catchTrialResponse\_questions\_time3: the answer to the catch trial

-catchTrialResponse\_questions\_time3\_corr: whether the catch trial is correct

## 

## Catch Trials/attention checks

Attention Check: In block 1 4 attention check were included. In 2 of these trials participants observed a statement (which was not drawn from the 60 statements list) and subsequently were presented with a list of 3 statements which included the one they previously saw. Their task was to indicate the previously seen statement. In the other 2 trials, after the presentation of the statement (which was not drawn from the 60 statements list), participants were asked to answer a question about the statement they had just seen.

In block 2 and 3, 6 attention check trials were inserted. 4 of these trials were identical to check trials in block 1. In the other 2 trials, instead of indicating their sharing decision or accuracy judgment, subjects were instructed to select a specific rating (for example: Select Definitely False).

In the raw data it is possible to identify these trials as their sentence number is 999. Note that these trials are not included in the analysis.

# 

# Data processing

The analysis process was identical for both experiments. All raw data were first cleaned, using Matlab scripts (see .m files in folder Exp1 🡪 RawData). The resulting clean data files are in folders Exp1 and Exp2. These were used for the main statistical analyses, performed in R and SPSS.

Here, we first describe the exact steps in the data cleaning procedure and each variable name as can be found in the processed data files.

## Cleaning and preparing the data

### Exp1

Raw data is divided into three folders. Raw data N is 162. Before running the Matlab script to clean the data, 2 subjects have been removed (one because the data have been saved incorrectly and one because restarted the task). Total N is 160.

### Exp 2

Raw data is divided into three folders. Raw data N is 102. Before running the Matlab script to clean the data, 2 subjects have been removed (because they completed the first block twice). Total N is 160. One subject completed the last block twice, data from the second attempt has been eliminated.

### Running the Matlab Script

Some participants completed the perceived accuracy block first, while some completed the sharing block first. The corresponding Matlab script need to be run for each data set to clean the raw data. Each script first load all three files for each data folder.

Data on each trial are saved in multiple rows for each participant. The Matlab script first keeps the last row for each trial, where the data from that trial is stored.

In the raw data it is possible to identify attention check trials, as their sentence number is 999.These trials are saved in another table and eliminated from the data.

For each subject, repeated information is determined by looking at which statements the subject saw in block 1. Information about the actual accuracy of each statement is also added.

Separately for perceived accuracy and sharing intentions, the mean accuracy and sharing rating is calculated for repeated true, repeated false, new true and new false statements.

Data at trial level is stored in a separate table. Data is sorted according to ID and sentence number.

Data from the attention check trials is properly analysed and stored in a new table.

Data at participant, data at trial level and data from the attention trials is saved in the same folder.

# Variable descriptions

## Data Trials

The files containing data at trial level have the following columns:

* PercAccuracy: perceived accuracy rating of each statement for each trial from
* Sharing: sharing rating of each statement for each trial
* Rep1: whether participants saw this sentence in block 1. 1 is repeated statement.
* ParticipantPublicID. For privacy reason, numbers from 1 to max N have been use to indicate the subject number.
* ShareFirst1: indicate which block participants completed first. 1 means that they completed the sharing block first.
* True1: Real accuracy of the statement. 1 means the statement is true
* SM: social media questionnaire score obtained by averaging the social media scaled scores.

## Subject Level Data

* Corr\_ShareAcc: correlation coefficient for each subj between accuracy and sharing
* Corr\_New: correlation coefficient for each subj between accuracy and sharing for new statements.
* Corr\_Rep: correlation coefficient for each subj between accuracy and sharing for repeated statements.
* Share\_New: mean sharing rating for statements not seen in block1
* Share\_Repeated: mean sharing rating for statements seen in block1
* Accuracy\_New: mean perceived accuracy rating for statements not seen in block1
* Accuracy\_Repeated: mean perceived accuracy for statements seen in block1
* ShareFirst1: indicate which block participants completed first. 1 means that they completed the sharing block first.
* ShareTrue: mean sharing rating of true statements
* Share False: mean sharing rating of false statements
* AccTrue: mean perceived accuracy rating of true statements
* Acc False: mean perceived accuracy rating of false statements
* Sh\_RepTrue: mean sharing rating of true repeated statements
* Sh\_RepFalse: mean sharing rating of false repeated statements
* Sh\_NewTrue: mean sharing rating of true new statements
* Sh\_NewFalse: mean sharing rating of false new statements
* Ac\_RepTrue: mean perceived accuracy rating of true repeated statements
* Ac\_RepFalse: mean perceived accuracy rating of false repeated statements
* Ac\_NewTrue: mean perceived accuracy rating of true new statements
* Ac\_NewFalse: mean perceived accuracy rating of false new statements

# Data Analysis

We used R to conduct the for Mediation Analysis and SPSS to conduct ANOVA and Linear Mixed Model analyses. When running the analysis on true information only, data of one subject in Exp1 (for the perceived accuracy model) and seven subjects in Exp2 (for the sharing model) caused the final Hessian matrix not to be positive definite, producing inflated degrees of freedom, due to lack of variability in their data. These subjects were thus removed from those specific analysis. Data file for these analyses are “DataTrialsLevelResultsExp1\_TRUE\_minus1” and “ResTrialLevel\_Exp2\_TRUE\_minus7”.

All R scripts and SPSS syntax to perform these analyses can be found on GitHub: For the ANOVA, we used repetition (new/ repeated) and ground truth (true/ false) as within subject variables, and perceived accuracy as the dependent variable. The same ANOVA was conducted for sharing ratings.