

Codebook

Markus D. Steiner, Florian I. Seitz, and Renato Frey

7/15/2019

Codebook Study 1

Data files included in data/study_1

- aspects.csv
- contents_models.RData
- participants_anonymous.csv
- ratings.csv
- study_long.csv
- study.csv

Description of data files and variables included therein

aspects.csv

A raw data file that contains the listed aspects, as well as the ratings of the strength of evidence and the sources and contents of each aspect. If categories are provided, the levels the variable can take in the data set are given in parentheses.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
soep_item	Which SOEP risk item was presented. Always “gen” for general.
aspect_ind	An index of the aspects listed by per participant. Starting at zero.
aspect	The text of the reported aspects.
r_risk	Answer to “How strongly does your description above support that you seek risks vs. that you avoid risks?” That is, the strength of evidence ratings, ranging from -50 to 50.
r_frequency	The answer to “How frequently do you normally experience or do what you described above?”. Categories were “Once per day” (day), “Once per week” (week), “Once per month” (month), “Once per year” (year), “Less than once per year, but on a regular basis” (less_reg), “Less than once per year, only once or a few times so far” (less_once), and “Not applicable” (NA).
r_social	The answer to “Does your description above include a comparison with another person?”. Categories were “Yes, it includes a comparison with another person” (1), “No, it does not include a comparison with another person” (0), and “Not applicable” (NA).
r_situation	The answer to “Does your description above include a personal experience?”. Categories were “Yes, it includes a personal experience” (1), “No, it does not include a personal experience” (0), and “Not applicable” (NA).

Variable	Description
r_control	The answer to “Are the outcomes and consequences of what you described above controllable or uncontrollable for you?”. Categories were “Rather controllable” (controllable), “Rather uncontrollable” (uncontrollable), and “Not applicable” (NA).
r_active	The answer to “Does what you described above include something you actively chose or something you passively experienced?”. Categories were “Active choice” (active), “Passive experience” (passive), and “Not applicable” (NA).

contents_models.RData

An RData file containing the outputs of several mixed effects models that are fitted in `r/study_1/2_study_1_analyses.R` on lines 951 to 996.

participants_anonymous.csv

A raw data file that contains the socio-demographic variables as well as the some control variables. If categories are provided, the levels the variable can take in the data set are given in parentheses. Lines containing **NA** stem from participants who did not complete the study.

Variable	Description
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
age	Reported age in years.
sex	Reported sex (female , or male).
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Montly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).
imc1	Was the first attention check answered correctly (correct , incorrect).
imc2	Was the second attention check answered correctly (correct , incorrect).
qual_focused	Rating of how focused participants worked on the study. Ranges from 0 to 100.
qual_effort	Rating of how motivated participants were while working on the study. Ranges from 0 to 100.
done	Whether the study was completed (1 , NA).
device	Device on which participants completed the study (desktop , laptop , tablet , mobile , other).
duration	Seconds needed to complete the study.

ratings.csv

A raw data file that contains the risk preference ratings.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.

Variable	Description
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
soep_item	Which SOEP risk item was presented. Always “gen” for general.
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.

study_long.csv

A pre-processed data file that contains various variables on an aspect level (i.e., one line for every aspect) .

Variable	Description
partid	The participant id.
aspect_ind	An index of the aspects listed by per participant. Starting at one.
sentiment	The sentiment of the aspects, obtained with the “bing” lexicon.
age	Reported age in years.
sex	Reported sex (female , or male).
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Monthly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).
qual_focused	Rating of how focused participants worked on the study. Ranges from 0 to 100.
qual_effort	Rating of how motivated participants were while working on the study. Ranges from 0 to 100.
duration	Seconds needed to complete the study.
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
aspect	The text of the reported aspects.
r_risk	Answer to “How strongly does your description above support that you seek risks vs. that you avoid risks?” That is, the strength of evidence ratings, ranging from -50 to 50.
r_frequency	The answer to “How frequently do you normally experience or do what you described above?”. Categories were “Once per day” (day), “Once per week” (week), “Once per month” (month), “Once per year” (year), “Less than once per year, but on a regular basis” (less_reg), “Less than once per year, only once or a few times so far” (less_once), and “Not applicable” (NA).
r_social	The answer to “Does your description above include a comparison with another person?”. Categories were “Yes, it includes a comparison with another person” (1), “No, it does not include a comparison with another person” (0), and “Not applicable” (NA).
r_situation	The answer to “Does your description above include a personal experience?”. Categories were “Yes, it includes a personal experience” (1), “No, it does not include a personal experience” (0), and “Not applicable” (NA).
r_control	The answer to “Are the outcomes and consequences of what you described above controllable or uncontrollable for you?”. Categories were “Rather controllable” (controllable), “Rather uncontrollable” (uncontrollable), and “Not applicable” (NA).
r_active	The answer to “Does what you described above include something you actively chose or something you passively experienced?”. Categories were “Active choice” (active), “Passive experience” (passive), and “Not applicable” (NA).
aspect_ind_cat	Categories of aspect index one to six (Aspect #1 , Aspect #2 , Aspect #3 , Aspect #4 , Aspect #5 , Aspect #6 , NA).
r_risk_binary	Indicator of whether it’s a pro- or contra-aspect (Pro , Contra).
n_aspects	The number of aspects a participant has listed.
m_risk	The average strength of evidence rating of a participant.
p_social	Proportion of Yes (1) answers to the r_social item.

Variable	Description
p_situation	Proportion of Yes (1) answers to the r_situation item.
p_control	Proportion of Rather controllable (controllable) answers to the r_control item.
p_active	Proportion of Active (active) answers to the r_active item.
rank_frequency	The rank of the frequency item (rank 1 is day , rank 6 is less_once) item.
med_frequency	Median of rank_frequency per participant.
SMRD	Standardized median rank difference. Ranges from -1 to 1.
mean_sentiment	The mean sentiment of an aspect per participant.
max_sentiment	The maximum sentiment of an aspect per participant.
min_sentiment	The minimum sentiment of an aspect per participant.

study.csv

A pre-processed data file that contains various variables on the participant level (i.e., one line for every participant).

Variable	Description
partid	The participant id.
age	Reported age in years.
sex	Reported sex (female , or male).
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
n_seeking	The number of pro-aspects a participant has listed.
n_avoiding	The number of contra-aspects a participant has listed.
m_risk	The average strength of evidence rating of a participant.
p_control	Proportion of Rather controllable (controllable) answers to the r_control item.
p_social	Proportion of Yes (1) answers to the r_social item.
p_situation	Proportion of Yes (1) answers to the r_situation item.
med_frequency	Median of rank_frequency per participant.
p_active	Proportion of Active (active) answers to the r_active item.
mean_sentiment	The mean sentiment of an aspect per participant.
min_sentiment	The minimum sentiment of an aspect per participant.
max_sentiment	The maximum sentiment of an aspect per participant.
SMRD	Standardized median rank difference. Ranges from -1 to 1.
aspect_ind	The number of aspects a participant has listed.
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Monthly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).
n_aspects	The number of aspects a participant has listed.
duration	Seconds needed to complete the study.

Codebook Study 2

Data files included in data/study_2

- aspects.csv
- contents_models.RData
- participants_anonymous.csv
- ratings.csv
- study_joint.csv
- study_long.csv
- study.csv

Description of data files and variables included therein

aspects.csv

A raw data file that contains the listed aspects, as well as the ratings of the strength of evidence and the sources and contents of each aspect. If categories are provided, the levels the variable can take in the data set are given in parentheses.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
soep_item	Which SOEP risk item was presented. Always “gen” for general.
aspect_ind	An index of the aspects listed by per participant. Starting at zero.
aspect	The text of the reported aspects.
r_risk	Answer to “How strongly does your description above support that you seek risks vs. that you avoid risks?” That is, the strength of evidence ratings, ranging from -50 to 50.
r_frequency	The answer to “How frequently do you normally experience or do what you described above?”. Categories were “Once per day” (day), “Once per week” (week), “Once per month” (month), “Once per year” (year), “Less than once per year, but on a regular basis” (less_reg), “Less than once per year, only once or a few times so far” (less_once), and “Not applicable” (NA).
r_social	The answer to “Does your description above include a comparison with another person?”. Categories were “Yes, it includes a comparison with another person” (1), “No, it does not include a comparison with another person” (0), and “Not applicable” (NA).
r_situation	The answer to “Does your description above include a personal experience?”. Categories were “Yes, it includes a personal experience” (1), “No, it does not include a personal experience” (0), and “Not applicable” (NA).
r_control	The answer to “Are the outcomes and consequences of what you described above controllable or uncontrollable for you?”. Categories were “Rather controllable” (controllable), “Rather uncontrollable” (uncontrollable), and “Not applicable” (NA).
r_active	The answer to “Does what you described above include something you actively chose or something you passively experienced?”. Categories were “Active choice” (active), “Passive experience” (passive), and “Not applicable” (NA).

contents_models.RData

An RData file containing the outputs of several mixed effects models that are fitted in `r/study_2/3_study_2_analyses.R` on lines 971 to 1016.

participants_anonymous.csv

A raw data file that contains the socio-demographic variables as well as the some control variables. If categories are provided, the levels the variable can take in the data set are given in parentheses. Lines containing **NA** stem from participants who did not complete the study.

Variable	Description
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
age	Reported age in years.
sex	Reported sex (female , or male).
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Monthly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).
imc1	Was the first attention check answered correctly (correct , incorrect).
imc2	Was the second attention check answered correctly (correct , incorrect).
qual_focused	Rating of how focused participants worked on the study. Ranges from 0 to 100.
qual_effort	Rating of how motivated participants were while working on the study. Ranges from 0 to 100.
done	Whether the study was completed (1, NA).
device	Device on which participants completed the study (desktop , laptop , tablet , mobile , other).
remembered	How well participants remembered what aspects they had listed in study 1. Ranges from 0 to 100.
similarity	How similar participants thought the aspects they had listed in study 1 were to those listed in study 2. Ranges from 0 to 100.
duration	Seconds needed to complete the study.

ratings.csv

A raw data file that contains the risk preference ratings.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The participant id.
cond_order	The order in which aspects and the risk preference rating were collected. Always “asp_judg”, meaning that always aspects were collected first and judgments were collected second.
cond_soepframe	The framing of the risk preference item. Always “std” for standard frame.
soep_item	Which SOEP risk item was presented. Always “gen” for general.
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.

study__joint.csv

A pre-processed data file that contains various variables of both studies on a participant level (i.e., one line for every aspect). “_s2” indicates data of Study 2, “_s1” indicates data of Study 1.

Variable	Description
partid	The participant id.
rating_s2	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
n_seeking_s2	The number of pro-aspects a participant has listed.
n_avoiding_s2	The number of contra-aspects a participant has listed.
m_risk_s2	The average strength of evidence rating of a participant.

Variable	Description
p_control_s2	Proportion of Rather controllable (controllable) answers to the r_control item.
p_social_s2	Proportion of Yes (1) answers to the r_social item.
p_situation_s2	Proportion of Yes (1) answers to the r_situation item.
med_frequency_s2	Median of rank_frequency per participant.
p_active_s2	Proportion of Active (active) answers to the r_active item.
mean_sentiment_s2	The mean sentiment of an aspect per participant.
n_aspects_s2	The number of aspects a participant has listed.
remembered	How well participants remembered what aspects they had listed in study 1. Ranges from 0 to 100.
similarity	How similar participants thought the aspects they had listed in study.
rating_s1	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
n_seeking_s1	The number of pro-aspects a participant has listed.
n_avoiding_s1	The number of contra-aspects a participant has listed.
m_risk_s1	The average strength of evidence rating of a participant.
p_control_s1	Proportion of Rather controllable (controllable) answers to the r_control item.
p_social_s1	Proportion of Yes (1) answers to the r_social item.
p_situation_s1	Proportion of Yes (1) answers to the r_situation item.
med_frequency_s1	Median of rank_frequency per participant.
p_active_s1	Proportion of Active (active) answers to the r_active item.
mean_sentiment_s1	The mean sentiment of an aspect per participant.
n_aspects_s1	The number of aspects a participant has listed.
pred_VUM_s1	Aggregated strength of evidence predicted by the value updating model (VUM).
pred_VUM_s2	Aggregated strength of evidence predicted by the value updating model (VUM).
m_similarity_between	Average similarity of aspects across studies.
m_similarity_within_s1	Average similarity of aspects within Study 1.
m_similarity_within_s2	Average similarity of aspects within Study 2.
p_overlap_between	Proportion of overlap of aspects across studies.
p_overlap_within_s1	Proportion of overlap of aspects within Study 1.
p_overlap_within_s2	Proportion of overlap of aspects within Study 2.
delta_rating	Absolute difference of the risk preference rating (SOEP general risk item) between study 1 and study 2.
delta_VUM	Absolute difference between “pred_VUM_s1” and “pred_VUM_s2”.

study_long.csv

A pre-processed data file that contains various variables on an aspect level (i.e., one line for every aspect) .

Variable	Description
partid	The participant id.
aspect_ind	An index of the aspects listed by per participant. Starting at one.
sentiment	The sentiment of the aspects, obtained with the “bing” lexicon.
age	Reported age in years.
sex	Reported sex (female , or male).
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Monthly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).

Variable	Description
qual_focused	Rating of how focused participants worked on the study. Ranges from 0 to 100.
qual_effort	Rating of how motivated participants were while working on the study. Ranges from 0 to 100.
remembered	How well participants remembered what aspects they had listed in study 1. Ranges from 0 to 100.
similarity	How similar participants thought the aspects they had listed in study 1 were to those listed in study 2. Ranges from 0 to 100.
duration	Seconds needed to complete the study.
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
aspect	The text of the reported aspects.
r_risk	Answer to “How strongly does your description above support that you seek risks vs. that you avoid risks?” That is, the strength of evidence ratings, ranging from -50 to 50.
r_frequency	The answer to “How frequently do you normally experience or do what you described above?”. Categories were “Once per day” (day), “Once per week” (week), “Once per month” (month), “Once per year” (year), “Less than once per year, but on a regular basis” (less_reg), “Less than once per year, only once or a few times so far” (less_once), and “Not applicable” (NA).
r_social	The answer to “Does your description above include a comparison with another person?”. Categories were “Yes, it includes a comparison with another person” (1), “No, it does not include a comparison with another person” (0), and “Not applicable” (NA).
r_situation	The answer to “Does your description above include a personal experience?”. Categories were “Yes, it includes a personal experience” (1), “No, it does not include a personal experience” (0), and “Not applicable” (NA).
r_control	The answer to “Are the outcomes and consequences of what you described above controllable or uncontrollable for you?”. Categories were “Rather controllable” (controllable), “Rather uncontrollable” (uncontrollable), and “Not applicable” (NA).
r_active	The answer to “Does what you described above include something you actively chose or something you passively experienced?”. Categories were “Active choice” (active), “Passive experience” (passive), and “Not applicable” (NA).
aspect_ind_cat	Categories of aspect index one to six (Aspect #1, Aspect #2, Aspect #3, Aspect #4, Aspect #5, Aspect #6, NA).
r_risk_binary	Indicator of whether it’s a pro- or contra-aspect (Pro, Contra).
n_aspects	The number of aspects a participant has listed.
m_risk	The average strength of evidence rating of a participant.
p_social	Proportion of Yes (1) answers to the r_social item.
p_situation	Proportion of Yes (1) answers to the r_situation item.
p_control	Proportion of Rather controllable (controllable) answers to the r_control item.
p_active	Proportion of Active (active) answers to the r_active item.
rank_frequency	The rank of the frequency item (rank 1 is day , rank 6 is less_once) item.
med_frequency	Median of rank_frequency per participant.
SMRD	Standardized median rank difference. Ranges from -1 to 1.
mean_sentiment	The mean sentiment of an aspect per participant.
max_sentiment	The maximum sentiment of an aspect per participant.
min_sentiment	The minimum sentiment of an aspect per participant.

study.csv

A pre-processed data file that contains varous variables on the participant level (i.e., one line for every participant).

Variable	Description
partid	The participant id.
age	Reported age in years.
sex	Reported sex (female , or male).
rating	The rating of risk preference (SOEP general risk item). Ranges from 0 to 10.
n_seeking	The number of pro-aspects a participant has listed.
n_avoiding	The number of contra-aspects a participant has listed.
m_risk	The average strength of evidence rating of a participant.
p_control	Proportion of Rather controllable (controllable) answers to the r_control item.
p_social	Proportion of Yes (1) answers to the r_social item.
p_situation	Proportion of Yes (1) answers to the r_situation item.
med_frequency	Median of rank_frequency per participant.
p_active	Proportion of Active (active) answers to the r_active item.
mean_sentiment	The mean sentiment of an aspect per participant.
min_sentiment	The minimum sentiment of an aspect per participant.
max_sentiment	The maximum sentiment of an aspect per participant.
SMRD	Standardized median rank difference. Ranges from -1 to 1.
aspect_ind	The number of aspects a participant has listed.
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
income	Monthly income, categorical (0_1k, 1k_2k, 2k_3k, 3k_4k, 4k_5k, 5k_x).
n_aspects	The number of aspects a participant has listed.
remembered	How well participants remembered what aspects they had listed in study 1. Ranges from 0 to 100.
similarity	How similar participants thought the aspects they had listed in study 1 were to those listed in study 2. Ranges from 0 to 100.
duration	Seconds needed to complete the study.

Codebook Similarity Ratings

Data files included in data/similarities

- aspects.csv
- participants.csv
- ratings.csv
- similarity_ratings.RDS

Description of data files and variables included therein

aspects.csv

A raw data file that contains all aspect pairs of the aspects listed by participants of studies 1 and 2.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The id of the participants from which the aspects stem from.
aspect_id_1	Identifier of the first aspect that was presented to the raters (the position on the screen, left or right, was randomized).
aspect_id_2	Identifier of the second aspect that was presented to the raters (the position on the screen, left or right, was randomized).
aspect_1	Text of the first aspect.

Variable	Description
aspect_2	Text of the second aspect.

participants.csv

A raw data file that contains the socio-demographic variables as well as the some control variables of the raters. If categories are provided, the levels the variable can take in the data set are given in parentheses. Lines containing NA stem from participants who did not complete the study.

Variable	Description
id	An id variable used as unique identifier of each row.
ratid	The rater id.
ts_id	Not needed variable, stemming from reused study code.
session	A session identifier, to identify each rating session of raters.
start	Unix timestamp of starting time.
end	Unix timestamp of completion time.
duration	Time in seconds it took to complete all ratings.
age	Reported age in years.
sex	Reported sex (female , or male).
yearsofedu	Years of education completed.
job	Employment status (employed , selfemployed , unemployed , student , retired).
imcscreening	Was the attention check in the screening answered correctly (correct , incorrect).
imc1	Was the first attention check answered correctly (correct , incorrect).
imc2	Was the second attention check answered correctly (correct , incorrect).
imc3	Was the third attention check answered correctly (correct , incorrect).
imc4	Was the fourth attention check answered correctly (correct , incorrect).
imc5	Was the fifth attention check answered correctly (correct , incorrect).
imc6	Was the sixth attention check answered correctly (correct , incorrect).
imc7	Was the seventh attention check answered correctly (correct , incorrect).
feedback	Free text comments provided by the raters.
done	Whether the study was completed (1, NULL).
N_rating	Total number of ratings completed by a rater.
N_aspects	Total number of aspect pairs.
imc_counter	Variable needed to control the flow through the study.
break_start	Variable needed to control the flow through the study.
break_status	Variable needed to control the flow through the study.
qual_focused	Rating of how focused participants worked on the study. Ranges from 0 to 100.
qual_effort	Rating of how motivated participants were while working on the study. Ranges from 0 to 100.
do_imc	Variable needed to control the flow through the study.
screening_ok	Variable needed to control the flow through the study.
first	Variable needed to control the flow through the study.
rating_group	Set of aspects that was rated.

ratings.csv

A raw data file that contains the similarity ratings.

Variable	Description
id	An id variable used as unique identifier of each row.

Variable	Description
ratid	The rater id.
id_aspects	The id of the rated aspect pair, in the “id” variable of “aspects.csv”.
rating	The similarity rating. Ranges from zero to 5.

similarity_ratings.RDS

A pre-processed data file containing the similarity ratings in the wide format.

Variable	Description
id	An id variable used as unique identifier of each row.
partid	The id of the participants from which the aspects stem from.
aspect_id_1	Identifier of the first aspect that was presented to the raters (the position on the screen, left or right, was randomized).
aspect_id_2	Identifier of the second aspect that was presented to the raters (the position on the screen, left or right, was randomized).
aspect_1	Text of the first aspect.
aspect_2	Text of the second aspect.
rating_...	The similarity ratings of every rater (the “...” represents the different rater ids). Ranges from zero to 5, is NA for the raters who did not rate the respective aspect pair.
m_similarity	The mean of the similarity ratings of the three raters who rated the same aspect pair.