Understanding the Processes Underlying Societal Threats using Novel Cluster-based Methods

A Data Management Plan created using DMPonline.be

Creator: Kim De Roover

Affiliation: KU Leuven (KUL)

Funder: European Research Council (ERC)

Template: ERC DMP +

Principal Investigator: Kim De Roover

Project Administrator: Kim De Roover

Grant number / URL: 101040754

ID: 202939

Start date: 01-10-2022

End date: 30-09-2027

Project abstract:

Social scientists are eager to answer questions about relations between constructs like beliefs or values. For example, do values affect climate change beliefs? Do perceived threats predict political beliefs? Do risk perception and susceptibility to misinformation determine vaccine hesitancy? Polarized beliefs about climate, politics, and vaccination are a societal threat and it is important to study what drives them. Large-scale survey data is gathered to do so.

Using regression to answer the questions ignores that constructs are not directly observable, but measured by survey items containing measurement error (challenge 1). Not correcting for this causes the studied effects to be underestimated and conclusions to be misguided.

When many groups are involved – such as many countries in the European Social Survey – the underlying processes likely differ across groups. For example, drivers of climate change beliefs may differ for countries experiencing extreme weather. Group-specific or multilevel analyses result in numerous group-specific regression slopes or random effects, making it hard to find which regression effects are different or similar for which groups (challenge 2).

Across many groups, the constructs' measurement is often inequivalent or 'non-invariant', for example, due to translation (challenge 3). A measurement model indicates how items measure a construct and disregarding non-invariance in this model invalidates the comparison of effects among constructs (i.e., one may find differences that are actually due to non-invariance).

By tackling challenges 1-3, the proposed mixture multigroup structural equation modelling framework provides the tools to break new ground in understanding what drives constructs like polarized beliefs. A clustering finds subsets of groups with common processes. Flexible measurement models account for non-invariance so that the clustering focuses on the processes. I will implement the methods in freely available software.

Last modified: 21-11-2023

Understanding the Processes Underlying Societal Threats using Novel Cluster-based Methods
GDPR Record

GDPR record

Have you registered personal data processing activities for this project?

- Not applicable

Understanding the Processes Underlying Societal Threats using Novel Cluster-based Methods
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Understanding the Processes Underlying Societal Threats using Novel Cluster-based Methods
ERC DMP +

Project information

Project Acronym

PROCESSHETEROGENEITY

Project Number

101040754

Data summary

Summary

- **Simulated data** (i.e., synthetic data) **will be generated**: These data will be generated and analyzed by means of the open source software “R”. They can be stored by the researcher in .csv format. But the data can be exactly replicated by other researchers based on the syntax files for generating the data. These syntax files will also include the seeds used for the random number generators, to ensure exact replicability of the data. A metadata .rtf text file will describe how each syntax can be used to replicate the data and the analysis of the data, as well as the versions of the software that was used during the project (i.e., the version of R and of R-packages that were used). At the end of the project, upon publication, the syntax files (incl. the random seeds) and the metadata will be shared (without restrictions) on the Open Science Framework, which allows to generate a DOI. The data can then be re-used by other researchers under the “CC BY attribution 4.0 international” license. Sharing the syntax and metadata, rather than the actual data sets, significantly lowers the required storage space (to about 1 MB, or less, per publication), while not hampering the replicability.
- **Openly available survey data** from the European Social Survey (www.europeansocialsurvey.org) and World Values Survey (www.worldvaluessurvey.org) **will be re-analyzed.** At the end of the project, the syntax used for handling and analyzing the data will be shared (without restrictions) on the Open Science Framework (with a DOI), together with a metadata .rtf file, that includes where to find the data online and how to replicate the data analysis by means of the syntax, as well as the versions of the software used for the analysis. The data can then be re-analyzed by other researchers under the “CC BY attribution 4.0 international” license.

The R-code for the newly developed methods resulting from this project, will be shared via GitHub.

FAIR data

1. Making data findable

When storing the data in the online repository of the Open Science Framework, suitable keywords will be used to make the data more findable.
For the simulated data (shared syntax and metadata), suitable keywords can be, for instance, "structural equation modeling", "regression", "measurement invariance", "large-scale surveys". The shared syntax and metadata will have a DOI.
For the re-analysis of openly available data sets (shared syntax and metadata), suitable keywords will refer to the type of analysis (e.g., "structural equation modeling") as well as the substantive questions answered by the re-analysis (e.g., "determinants of climate change beliefs"). The shared syntax and metadata will have a DOI.

2. Making data openly accessible

- **For the simulated data**: The syntax used for generating (and analyzing) the data will be shared, including the seeds used for the random number generators, to ensure exact replicability of the data. A metadata .rtf file will describe how each syntax can be used to replicate the data and the analysis of the data, as well as the versions of the software that was used during the project (i.e., the version of R and of R-packages that were used). The metadata will also be shared. The syntax files (incl. the random seeds) and the metadata will be shared (without restrictions) on the Open Science Framework, with a DOI.
- **For the re-analyzed openly available survey data** from the European Social Survey (www.europeansocialsurvey.org) and World Values Survey (www.worldvaluessurvey.org): At the end of the project, the syntax used for handling and analyzing the data will be shared (without restrictions) on the Open Science Framework (with a DOI), together with a metadata .rtf file, that includes where to find the data online and how to replicate the data analysis by means of the syntax, as well as the versions of the software used for the analysis.

The repository of the Open Science Framework is a trusted repository that is the go-to repository for other researchers involved in psychometric and sociometric analysis (which is the relevant field for my project).
The R-code for the newly developed methods resulting from this project, will be shared via GitHub, which is also standard practice in the field.

3. Making data interoperable

The syntax files will be shared in .m format, which can be read and edited by any text editor and used directly in the open-source software R.
The metadata will be shared as a .rtf text file, which is openly accessible and facilitates long-term usability.

4. Increase data re-use

The shared data (i.e., syntax and metadata) will be shared under the “CC BY attribution 4.0 international” license, which means that researchers can use, share and adapt the data as long as they give appropriate credit, provide a link to the license, and indicate if changes were made.

The data will be stored and re-usable for at least 10 years after the end of the project.

All syntax files will include documentation at the top of the file (explaining the content of the data set, including definitions of the variables), as well as explanatory lines throughout the code (explaining the steps of the analysis and/or data generation).

The R-codes pertaining to the newly developed methods developed in this project will be shared under an open source software license, specifically, a permissive licence allowing reuse, modification and distribution.

5. Allocation of resources and data security

Since only the syntax files (including random seeds) and metadata are needed to exactly replicate the simulated data and data analysis, not a lot of storage space is needed for this project (i.e., 1 GB or less per publication).

In the repository of the Open Science Framework, data storage is free. In terms of storage space, the following restrictions apply: No single file can be larger than 5GB, and OSF storage is limited to 50GB for public (i.e., openly shared) projects. It is also a trusted repository, ensuring long-term preservation.

The responsibility for this data storage and sharing lies with the PhD-student working on the concerned subprojects, under the supervision of the principal investigator (dr. Kim De Roover).