AFFECTOMETRICS: IMPROVING THE MEASUREMENT OF AFFECT IN DAILY LIFE

A Data Management Plan created using DMPonline.be

Creators: Francis Tuerlinckx, Peter Kuppens, Eva Ceulemans, n.n. n.n., Kim De Roover, Kristof Meers

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n., Francis Tuerlinckx, Peter Kuppens, Eva Ceulemans, Kim De Roover

Data Manager: Kristof Meers

Grant number / URL: C14/23/062

ID: 205648

Start date: 10-01-2023

End date: 01-03-2027

Project abstract:

Momentary affective experiences in daily life are relevant for the prediction, diagnosis, and prevention of mental health problems. The current state of the art of measuring affect in daily life is the Experience Sampling Methodology (ESM): Using a smartphone app, participants self-report on their momentary affective state at random moments throughout the day. The aim of this project is to develop a systematic affectometric research program to improve the science of affect measurement in four pivotal ways: (1) investigating and improving the validity of affect measurements through unraveling the underlying response-generating processes, (2) improving the measurement of affect change and regulation through innovative data collection methods, (3) improving the measurement of dynamic affect structure by appropriately accounting for measurement quality and measurement models, and (4) investigating the added value of improved affect measures for the prediction of mental health and well-being.

Last modified: 20-03-2024

AFFECTOMETRICS: IMPROVING THE MEASUREMENT OF AFFECT IN DAILY LIFE

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

In the original project proposal, we have given a concise overview of the data used in this project (see Table 1). For convenience, this is repeated here.

WP	Objective	Scope	Data	Models
1	Validity of self-reported affect ratings	single item; single moment	simulations; lab/online; ESM	evidence accumulation models (drift diffusion model); data-driven modeling
2	Measuring affect change & regulation	single item; across time	simulations; ESM	autoregressive (AR) models; data driven modeling
3	Measuring dynamic affect structure	multiple items; across time	simulations; ESM	(vector) autoregressive models; structural equation models
4	Predicting with improved affect measurements	all the above; mental health	lab/online; ESM; questionnaires	data-driven modeling

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Experiments	responses and reaction time collected in the context of standardized laboratory tasks	N	D	N	.csv	<1GB	
Experience Sampling data 1	participants report on their current emotional/cognitive state, behavior, and context throughout their daily life by means of a custom-built app on their own smartphone or a research-dedicated smartphone	N	D	N; T	.csv	<1GB	
Experience Sampling data 2	participants report on their current emotional/cognitive state, behavior, and context throughout their daily life by means of a custom-built app on their own smartphone or a research-dedicated smartphone	E	D	N; T	.csv	<1GB	
Mobile sensing data	this involves harvesting data on smartphone use that is automatically recorded by the participant's smartphone including phone activity, GPS location, accelerometer activity, ambient light, app use, charge time, screen-on time, photo access, keyboard use, SMS and call frequency	E	D	Observational data; A, N, T	.json	<1TB	
Background data	baseline demographic information, individual differences data in terms of personality, mental health	N	D	N	.csv	<1GB	
Simulated data	data simulated under statistical models such as for instance the Drift Diffusion Model or Vector-Autoregressive models	N	D	N	.csv; .Rdat	<100GB	
Analysis syntax	Syntax associated with simulations or empirical analyses	N	D	T	.R; .txt; .Rmd;	<1GB	
Software	computer programs to collect and process data. This involves packages and shinyapps in R	N	D	SO	.txt	<1GB	
Preregistrations	The empirical and simulation studies will be preregistered	N	D	T	.txt; .docx; .pdf	<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Data set name	Source	Description	Persistent Identifier
Experience sampling data 2	EMOTE database	The EMOTE (Everyday Measures of Temporal Emotions) Database is an open-access, searchable, and cumulative database of experience sampling data on daily emotional functioning. Experience sampling methods (also known as ecological momentary assessment) involve sampling human experiences in real time (or close to it) across multiple measurement occasions. The EMOTE database contains data from 23 ESM studies which will be used as part of a secondary analysis project	https://emotedatabase.com/
Mobile sensing data	Niemeijer K, Mestdagh M, Verdonck S, Meers K, Kuppens P Combining Experience Sampling and Mobile Sensing for Digital Phenotyping With m-Path Sense: Performance Study JMIR Form Res 2023;7:e43296	We are referencing information that is detailed in the publication available as the actual datasets are not available for public sharing.	https://doi.org/10.2196/43296

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

Yes, the empirical data (hence not simulated data, software or analysis code, etc.) we will collect and/or analyze are data from human participants.

Dataset name	SMEC approval status
Experiments	G-2020-2772-R3(AMD)/G-2023-7257 (Experiments on the affective consequences of expectations)
Experience Sampling data 1	G-2021-4482-R3(AMD) (ESM study on careless responding) G-2022-5827-R3(AMD) (ESM study on measurement error) G-2022-4779-R3(AMD) (ESM study on affect intensity profiles) G-2021-2990-R2(AMD) (ESM study on the pros and cons of event-based sampling)
Experience Sampling data 2	SMEC approval will be obtained to reuse the data (see Compliance form E-2023-4083)
Mobile sensing data	SMEC approval will be obtained to reuse the data (see Compliance form E-2023-4083); the SMEC approval number of the original study is G-2020-2200-R3(AMD)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- Yes (Provide PRET G-number or EC S-number below)

Yes. The background data will consist of information on the age, gender, and mental health status of the participants. These background data (all or a subset of it) will be obtained in every empirical study; therefore in the table below the background data are not separately mentioned as they are part of other data sets. In addition, the (reused) mobile sensing data involves data on participants' personal smartphone use.

Dataset name	G number
Experiments	G-2020-2772-R3(AMD)
Experience Sampling data 1	G-2022-5827-R3(AMD)
Experience Sampling data 2	Privacy approval will be obtained (see Compliance form E-2023-4083)
Mobile sensing data	Privacy approval will be obtained to reuse the data (see Compliance form E-2023-4083); the approval number of the original study is G-2020-2200-R3(AMD)

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The datasets derived from participant data across all work packages are exclusively designated for academic and research purposes and are not earmarked for commercial exploitation. However, the technological advancements achieved (specifically regarding methodology to flag data points from ESM that are the result of careless responding or compliance (WP1), the affect intensity profile drawings in ESM and a framework for analyzing them (WP2) and the episode-contingent measurement burst framework (WP2)) can be used for commercial valorization. If the developed methodologies prove to be successful, we will get into contact with LRD in order to explore tech transfer options. Obviously, our preferred partner for this will be m-Path as it has been developed by our group and will spin off in Spring 2024. However, we will act transparently in order to avoid any potential conflict of interest. In any case, KU Leuven will be the IP holder for all methodology, data and software created in the project.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

By using data from the EMOTE database, we agree to their terms and conditions. This means that we do not own the data and do not have the right to share the data with third parties, though we may communicate the results of processing the data in scientific publications. We will comply with the data retention policy of the KU Leuven (10 years), this is allowed under the agreement with the EMOTE database.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

All data and associated files will be stored in a data repository directory. We will provide a folder structure that reflects the content. In addition we will provide readme and metadata files that explain the data structures and content of the files where needed. The information necessary for a secondary analyst to use the data accurately and effectively will be documented in a file (txt, csv, or word-file which contains, e.g., the info on dates, times and number of the ESM measurements; the variable list and legend; info on accessibility of the dataset; and info on processing operations on data files).

Will a metadata standard be used to make it easier to find and reuse the data?

If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

There is no formally acknowledged metadata standard specific to our discipline, so we will rely on the DDI standard (Data Documentation Initiative) for the description on the project level. All information will be provided in txt, csv, or word-file. Some pseudonomyzed data sets (e.g., the data from the lab experiments) will be shared through the Open Science Framework (OSF) where they will be stored on servers within the EU.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Other (specify below)
- Large Volume Storage

The data from this project will be stored on OneDrive for convenient access and on Large Volume Storage for long-term preservation. We will process the data exclusively on computers with encrypted hard drives. Data associated with publications may also be published on open data repositories such as OSF and the EMOTE database.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

Yes, we have 5 TB of space available and a budget if more would be necessary (but we do not anticipate that more data storage is needed).

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

To secure our data, we store it within KU Leuven's Large Volume Storage and OneDrive, both compliant with high and medium confidentiality standards, respectively. KU Leuven's ICT solutions adhere to stringent university-wide information security protocols. Access permissions for the raw network storage managed by the faculty's ICT service are strictly regulated, delegated, and audited by designated data managers who are trained for this role, regardless of their IT expertise. For the code files, which include pseudonyms, access is tightly controlled by a dedicated data manager, with the Principal Investigator serving
as an alternate overseer. Researchers involved in the project are granted access exclusively to pseudonymized data, ensuring that personal identifiers are not disclosed.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The anticipated costs for data storage and backup for the duration of the research project are estimated at €2378.5, based on a block 5TB storage for 5 years. These expenses can be covered using our current grant money and through future grant applications.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Large Volume Storage (longterm for large volumes)
- Other (specify below)

A data archive will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform KU Leuven policy. Some data will in addition also be available on the OSF and EMOTE servers (although we cannot guarantee a 10 years period).

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The anticipated costs for data storage and backup for the duration of the research project are estimated at €4757, based on a block 5TB storage for 10 years. These expenses can be covered using our current grant money and through future grant applications.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)
- Yes, as open data
- Yes, as embargoed data (temporary restriction)

For the empirical data that we collect during this project (ESM and experimental data), the pseudonymized versions will be made open, subject to an embargo period to secure the research team's time to fully analyze the data and publish the findings. This measure is to prevent premature dissemination of the data which could lead to potential intellectual property issues or misappropriation of the research concepts by third parties. The data will be shared using various ways (e.g., OSF and/or the EMOTE database). Some of the data will be only available upon approval (e.g., this holds for data in the EMOTE database).

Due to the sensitive nature of the secondary mobile sensing data and the fact that certain components cannot be effectively pseudonymized, these data in this project will not be made publicly available for external sharing. This ensures the protection of participant privacy and data confidentiality.

The preregistrations, simulation data, and syntax will be made publicly available on the open science framework.

If access is restricted, please specify who will be able to access the data and under what conditions.

Participants' personal information (e.g., contact information, names, etc.) will never be shared. The secondary mobile sensing data will not be made publicly available as these cannot be sufficiently pseudonymised.

Experimental, ESM and background data will be made available after a period of embargo. The data will be shared using various ways (e.g., OSF and/or the EMOTE database). Analysis syntax and code to perform simulations will be uploaded to the Open Science Framework and/or Gitlab. R packages and associated functions will be uploaded on CRAN and Gitlab under a GPL-3 license.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- Yes, intellectual property rights
- Yes, privacy aspects

Due to the sensitive nature of the secondary mobile sensing data and the fact that certain components cannot be effectively pseudonymized, these data in this project will not be made publicly available for external sharing.
Some of the intellectual output may have the potential for valorization. In that case, the KU Leuven IP should not be shared freely and without restriction.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

Experimental, ESM and background data will be made available after a period of embargo. The data will be shared using various ways (e.g., OSF and/or the EMOTE database). Analysis syntax and code to perform simulations will be uploaded to the Open Science Framework and/or Gitlab. R packages and associated functions will be uploaded on CRAN and Gitlab under a GPL-3 license.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- GNU GPL-3.0 (code)

Empirical data (experiments, ESM), simulation data, syntax, and preprints will be made publicly available on the Open Science Framework. When doing so, a CC-BY 4.0 license is
provided.
Any R packages that result from the project will be published under the GNU GPL-3.0 license.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository
- Yes, my dataset already has a PID

For the newly collected data, a permanent identifier is added to data upon deposit in a repository. For the existing data, PID are already in place (unless the data cannot be shared).

What are the expected costs for data sharing? How will these costs be covered?

Sharing data at the Open Science Framework, CRAN and Gitlab is free.

Responsibilities

Who will manage data documentation and metadata during the research project?

PhD Students and postdoc researchers, technicians, administrative personnel under the supervision of the PIs.

Who will manage data storage and backup during the research project?

KU Leuven, specifically the PIs of the project (Francis Tuerlinckx, Agnes Moors, Eva Ceulemans, Kim De Roover and Peter Kuppens).

Who will manage data preservation and sharing?

Francis Tuerlinckx, Agnes Moors, Eva Ceulemans, Kim De Roover and Peter Kuppens

Who will update and implement this DMP?

Francis Tuerlinckx, Agnes Moors, Eva Ceulemans, Kim De Roover and Peter Kuppens