Plan Overview

A Data Management Plan created using dmptool

Creator: Katharina Fellnhofer

Affiliation: Harvard University

Funder: European Commission (ec.europa.eu)

Template: Digital Curation Centre

ORCID iD: 0000-0002-8421-1547

Project abstract:
Because of the nature of entrepreneurship, it is postulated, in line with emotions-as-information theory, that intuition affects decision-making. How intuition affects profits and growth of female and male ventures as well as access to finance, which are crucial for survival, has not to date been objectively investigated. Applying the social-psychology theory of the stereotype content model, ROLLER-COASTER’s aim is to investigate entrepreneurs’ and venture capitalists’ (VCs) intuition and its impact on success. Three research objectives are targeted: 1) to create a solid interdisciplinary framework 2) to explore the ‘entrepreneurial gender ceiling’ 3) to enhance related theories and concepts Two study phases focus on achieving the objectives: Study 1: Does intuition among entrepreneurs facilitate profit and growth similarly for men and women? Based on Orbis data, self-reports, and behavioral data relating to at least 200 entrepreneurs of both sexes, it is proposed to analyze if female entrepreneurs show higher attributions to intuition but use it less effectively in realizing firm profit and growth than males do. Study 2: Does VCs’ intuition influence investment decision preferences in either female- or male-led ventures? Based on Study 1, pitches will be produced with switched genders and shown to at least 200 VCs to explore whether VCs with high attributions and use of intuition are more likely to invest in male than in female ventures. ROLLER-COASTER will support European educators and policy-makers as they develop instruments to support entrepreneurs --- in particular females as an under-exploited source of economic growth and jobs --- to use their intuition to make faster, more accurate, and more confident decisions. This project is being carried out by ETH Zurich (Chair of Entrepreneurial Risk) together with Harvard University (Department of Sociology - Harvard Sociology). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 882168. Responsible Principal Investigator is Dr. Katharina Fellnhofer, currently a senior researcher at the Chair of Entrepreneurial Risk at ETH Zurich. Katharina Fellnhofer is also an
Associate at the Department of Sociology - Harvard Sociology. The European Commission communicates the project as follows:
https://cordis.europa.eu/project/id/882168

Last modified: 05-18-2021

Copyright information:
The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.
Data Collection

What data will you collect or create?

Two research designs for two studies will be utilized, as detailed below. Both studies are incorporated in one survey and experiment. Thus, participants are engaged approximately 30 minutes in following studies:

Study 1. Based on the Stereotype Content Model,[1] I argue that the *attribute to intuition* and its effective use affect ventures’ success. For instance, high-performing SMEs are characterised by an entrepreneur’s high *attribute to* and effective use of intuition. Moreover, my assumption that women will show different levels of intuition than men, potentially leading to different degrees of success, is built on prior work related to career reasons[2] and a model of psychological biases for stereotype judgments that underlie female and male decision-making.[3]

In order to measure the independent variable entrepreneurial intuition via self-reports, I will use established (cognitive-style) scales[4] with analytic-intuitive metrics. For instance, I will rely on previous theoretical concepts suggesting predictors of *attributions to intuition* (e.g., perceived previous success, and intuitive cognitive style) and the effective use of intuition (e.g., experience, domain-relevant knowledge).[5] The dependent variable, **profit**, will be measured by archival data on net profit or loss and cash flow taken from Amadeus or Orbis, the industry-leading company database. Next, I will measure growth by estimating whether the SME grew in terms of revenue and employees. The control variables are type of individual (private investor, entrepreneur, non-entrepreneur, professional investor or venture capitalist (VC)), gender, education, nationality, age, entrepreneurial/investor knowledge, industry/job experience, risk attitudes, social background/perceptions, personal disposition and emotional sensibility.

The self-reported survey will be enriched with Experiment 1 to collect behavioural data as follows. As nonconscious emotional information can boost decision-making accuracy (effective use of intuition) and increase confidence (attribute to intuition) I will apply a novel technique,[8] adapted to the entrepreneurial setting, to a concurrent, emotion-free decision task designed to measure the level of intuition in which subliminal emotional information is presented to participants while they make conscious decisions. They will perform a non-random, real-life task that requires them to use intuition while being exposed to suppressed emotionally charged positive and negative images as stimuli (e.g. from the Open Affective Standardized Image Set (OASIS)).[7] The chosen images show equal mean arousal and equal mean valence between genders. To adapt this novel technique to the entrepreneurial setting I will use a modified version of the random-dot-motion task. This task has been extensively used to investigate decision-making.[8] Participants are asked to invest time or money based on profit growth of a company. Participants will be required to decipher the moving future direction of venture profit. To add an emotional factor, emotionally charged images – on a nonconscious level – will be flashed that correlate to the direction of the arrows in the venture profit figures. In line with Lufityanto et al.’s (2016) procedure (subjects will complete three blocks of 96 trials instead of 144) and findings I expect these positive and negative images to make participants more likely to choose the correct direction. Furthermore, the higher the success rate, the higher one’s effective use of intuition is deemed. Based on the population’s sample size calculations and to permit a comparative study, the first sample will consist of at least 200 entrepreneurs (100 female- and male-led ventures – half based in the USA and half in Europe – recruited by Qualtrics Research Support Service and https://www.prolific.co/); half will serve as control group 1 and will not be exposed to emotionally charged images during Experiment 1.

Referring to stereotyping, I argue that the *attribute to intuition* and its effective use affect VCs’ willingness to invest in (female- or male-led) SMEs. For instance, the greater a VC’s *attribute to* and use of intuition, the more likely that VC is to prefer to invest in male entrepreneurs. This investigation will methodologically enhance prior research that examined only investment angels’ intuitive decision-making criteria based on self-reports[9] and the intuitive decision-making of high- and low-performing financial traders.[10]

Methodologically, I will repeat both the self-reported survey and experiment with a sample of at least 200 financial professionals/VCs (from the USA and Europe with equal gender distribution, recruited by Qualtrics Research Support Service and https://www.prolific.co/), half of whom will serve as control group 2 and will not be exposed to emotionally charged images during the repeated experiment. I will explore VCs’ level of intuition in greater depth with an experiment as follows. An entrepreneurial pitch will be presented via (1) a written one-page no-name investment teaser (as induction for neutral emotional status), (2) a podcast recorded from (3) a one-minute-clip presented either by a female, a male or both entrepreneurs. Each participant will be randomly assigned to one of three groups: treatment group 1 (pitches presented by female), treatment group 2 (pitches presented by male) or control group 3 (pitches presented by both). Immediately after each (written, audio or video) pitch, the participant has to assess his or her willingness to invest. The dependent variable willingness to invest will
be measured with intention scales in line with prior related investigations. [11] Thereby, the participant will be requested to reply to statements regarding what proportion of (item 1) their own savings and (item 2) others' savings they would be willing to invest in the enterprise and to what extent they would (item 3) provide further support and devote other resources to the SME. Moreover, they will be asked to estimate (item 4) the previous financial success of the observed business from the pitch and (item 5) how likely it is that the business would be an overall future success in terms of profits and growth. Finally (item 6), willingness to invest will be measured by whether the observer recommends due diligence. They can also invest virtual money.

Five control groups. Half of the participants will not be exposed to emotionally charged images during Experiment 1. Control group 3 during Experiment 2. Control group 4 will comprise a set of at least 200 non-entrepreneurs. The control group 5 to simultaneously collect evidence from mouse tracking to get additional information.

Details about the collected data:

- Variables collected via Amadeus:

  Firm's name, average ROE using P/L before tax, location, NACE code, gender of entrepreneur, revenue from year 2010 to 2019, profit from 2010 to 2019, cash flow from 2010 to 2019, staff from 2010 to 2019. In principle we will analyze all available data offered by Amadeus or Orbis related to the 16 companies which are implemented in our experiment.

- Variables collected via Qualtrics survey (excerpt):

  Stakeholder group (entrepreneur, financiers, professional investors, non-entrepreneurs), age, gender, nationality, average yearly net income over the last three years (rounded in thousand), highest educational degree, investment experience, entrepreneurial experience, working experience overall, industry experience with high-tech and low tech sectors. Several scales and items related to risk awareness, intuitive/rational decision making, valence and arousal to (subliminal) images and overall life satisfaction. After reading an investment teaser, and watching an audio and video podcast of an investment opportunity the participant is asked to assess his/her willingness to invest. At the end of the survey the participant will be redirected to the Pavlovio platform where the PsychoPy experiment will be performed. There will be an ID collected which is anonymous to connect the Qualtrics survey data to the PsychoPy experiment data because those are two different platforms.

- Variables collected via PsychoPy (excerpt):

  intro_mouse.x, intro_mouse.y, intro_mouse.leftButton, intro_mouse.midButton, intro_mouse.rightButton, intro_mouse_time, intro_mouse_clicked_name, answer_invest_money.x, answer_invest_money.y, answer_invest_money.leftButton, answer_invest_money.midButton, answer_invest_money.rightButton, answer_invest_money_time, answer_invest_money_clicked_name, answer_invest_time.x, answer_invest_time.y, answer_invest_time_leftButton, answer_invest_time_midButton, answer_invest_time_rightButton, answer_invest_time_time, answer_invest_time_clicked_name, answer_invest_confidence.x, answer_invest_confidence.y, answer_invest_confidence_leftButton, answer_invest_confidence_midButton, answer_invest_confidence_rightButton, answer_invest_confidence_time, answer_invest_invested_money, invested_time, confidence, answer_happy.x, answer_happy.y, answer_happy_leftButton, answer_happy_midButton, answer_happy_rightButton, answer_invested_time, answer_invested_money, invested_confidence, answer_happy clicked_name, trials_thisRepN, trials_thisTrialN, trials_thisN, trials_thisIndex, trials.run, Graph, Stimulus, profit_growth, GraphPath, StimulusPath, participant, mobile, date, expName, psychopyVersion, OS, frameRate, control group, positive feedback, block_thisRepN, block thisTrialN, block thisN, block thisIndex, block run, rate_of_return, block_sequence_thisRepN, block_sequence_thisTrialN, block_sequence_thisN, block_sequence_thisIndex, block_sequence.run, loop_nr, condFile, biometrics such as eye tracking at different stages during the experiment for attention checks (not for identification purposes). Time and mouse tracking. Eye tracking if participant agrees explicitly and technical feasible from participant's system.

References

How will the data be collected or created?

STANDARDS AND METHODS
This research project takes advantage of common standards in social science research and well-accepted methodologies as follows. The data collection phase will be divided into a pre-testing phase with individuals from three different target groups (25 participants from each group: entrepreneurs, financiers, non-entrepreneurs), and the global data collection phase (rest from each group to reach at least 600 participants).

A web-based survey (via Qualtrics) will enable the collection of wide-ranging data. Given the nature of the project, while individuals’ gender, age, experience, nationality and educational background will be gathered, average income of the last three years in thousand USD rounded no further personal information about the participants will be collected or attempted to be determined. This research project has no interest in direct personal information. Data will be collected in both Europe and the USA.

For the production of three pitch clips, the engaged female (I as PI) and male entrepreneurs signed a separate consent form that permits the clips to be shared and distributed for research purposes. A Harvard template has been used.

The data will be collected separately in individual excel files at each platform (Qualtrics for the survey, https://www.prolific.co/ for data collection and Pavlovia/PsychoPy for the experiments). The files are saved on those platforms. After the study phase has been closed all individual files per participants will be merged in one global Excel and SPSS file. Those two files will be saved locally on two laptops owned by the PI/myself, in a OneDrive cloud and on one external hardware device.

QUALITY MANAGEMENT
Research assistants and/or students will independently perform quality assurance analyses and checks of the data during the data collection phase (e.g., Harvard Statistics Consulting Service | Department of Statistics). While all of the different versions will have been handled by the PI, closing quality and consistency checks of the final data will be also performed by additional senior researchers (e.g, Harvard Statistics support service).

MANAGEMENT OF DIFFERENT VERSIONS
During and after the global collection phase, relevant information and data will be prepared for a diverse set of publications. Data will be downloaded from the online survey tool and Pavlovia and stored with a date stamp.

Documentation and Metadata

What documentation and metadata will accompany the data?

WHICH INFORMATION IS NEEDED TO READ THE DATA?
The data are collected in individual SPSS and Excel files and therefore saved separately as SPSS data or syntax files and Excel files and will be merged in one final Excel and SPSS file. Thus, all relevant data can be read using Microsoft Excel. The analysed data will be also stored in SPSS files. The files are saved locally on the two PI’s laptops, in a OneDrive cloud and on one external hardware devices. For the analysis phase, all data will be assembled in one SPSS file.
and one Excel file. Overall, no personally specific information is needed when reading or interpreting the data. SPSS and Microsoft Excel are sufficient to carry out all data collection and analysis tasks.

WHERE STANDARDS WILL BE USED FOR READING DATA?
Materials will be transcribed into Microsoft Office applications such as Word, Excel and PowerPoint or SPSS (data or syntax) files. For the objective of long-term open access publications (e.g., accepted peer-reviewed manuscripts of the research results), related documents will be converted into PDF files. The publications will outline the data collection processes and all details needed to repeat the process. As a result, no special knowledge or information is required for reading or interpreting the data in the future. This data management plan and the peer-reviewed publications of the research results will provide all of the necessary information to follow and repeat the project. This documentation will be freely available through the research platform and/or open access academic publications to be published.

METADATA
The metadata will be published (e.g. OpenAIRE). Open access publications will at least include details on the applied methodology, reliability and validity analyses, analytical and procedural information, definitions of the chosen items, variables of previous contributions upon which it is built, units of measurement and assumptions made. These types of documentation will help those reviewing the data to understand and repeat the research project if necessary.

ETHICS AND LEGAL COMPLIANCE

How will you manage any ethical issues?

SECURITY STANDARDS TO PROTECT ALL DATA
A carefully crafted plan has been devised to manage ethical issues. First, before obtaining access to the full web-based questionnaire and experiment (which have access secured through SSL certificates), the potential participant will be introduced to and informed about the collected data via an introduction page (see Questionnaire and Consent Form). All communications with Pavlovia require a secure HTTPS connection and are encrypted using secure ciphers. The necessary approvals by Ethical Boards of ETH Zürich and Harvard University are collected in advance (e.g., Committee on the Use of Human Subjects, a University-Area Institutional Review Board at Harvard University).

Overall, we will go through our data collection strategy with all the necessary ethical boards to ensure alignment with all ethical guidelines on a global basis. Ethics is of the highest priority. Personal data will not be transferred to third parties for any purpose. The results of the project and all intellectual property rights (IPR) related to all results are owned by the researcher and beneficiaries in accordance with the rules of the funding agency, the Grant Agreement and signed Partnership Agreement between Harvard University and ETH Zürich. In addition, beside further research and educational purposes no reuse and data sharing or further exploitation of the data set are planned at this stage. As a consequence, permissions to reuse the generated data or future data sharing will be based on these parameters.

Furthermore, we will follow The European Code of Conduct for Research Integrity of ALLEA (All European Academies).
The PI has conducted relevant training (e.g., Harvard Training dedicated to Responsible Conduct of Research, see Data Security Submission).

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

INTELLECTUAL PROPERTY RIGHTS
The results of the project and all intellectual property rights (IPR) related to all results are owned by the researcher in accordance with the rules of the funding agency, the Grant Agreement and signed Partnership Agreement between Harvard University and ETH Zürich. The researcher and beneficiary will own the copyright and IPR of any data that will be collected or created throughout the research project. In addition, beside further research and educational purposes no reuse and data sharing or further exploitation of the data set are planned at this stage. Throughout the research project, no other third-party data (with exclusion of Amadeus/Orbis data) will be used. As a consequence, permissions to reuse the generated data or future data sharing will be based on these parameters.

LICENCES
We aim for open access articles to be distributed under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

EXTERNAL DATA
We will use external data from Orbis and Amadeus that collect venture data related to profit and growth. These data are public available via a licence.

Storage and Backup

How will the data be stored and backed up during the research?

DATA SAVING MANAGEMENT
The data will be stored as long as necessary for the analysis and dissemination of research results via open access. Backups will take place on a regular basis and at different storage sites. All the collected data will be inputted and stored in widely available formats (e.g., Excel, Word, PDF and SPSS) on two internal laptops (500 GB each), one OneDrive cloud server (4 TB) and one external hard drive owned by the researcher (4 TB). All the questionnaire structures will be saved as word document from Qualtrics. The PI will be responsible for backup and recovery during the research project on a regular basis and at the final stage of the project.

How will you manage access and security?

DATA SECURITY
When information is transferred over the Internet, complete confidentiality cannot be guaranteed. The host of the system collecting the data may collect further information without our knowledge and make that information available to others. Alternative methods such as an e-mail or paper-based questionnaire might decrease anonymity, although confidentiality will be maintained. However, the method chosen for the research purpose appears to be optimal with respect to the objectives. Survey participants are informed in advance of these circumstances. Overall, the data gathered will not be shared with anybody expect researchers, collaborators, and funding bodies (if necessary).

The three main properties of data security are availability (ensuring data is not lost or inaccessible), integrity (not corrupted or maliciously manipulated) and privacy (not disclosed or used improperly). To follow data security, data controllers have been appointed. There are one Data Protection Officer appointed by each host organization. People who volunteer to participate as subjects in research do so with the understanding that the researcher(s) will protect their identity and the information obtained about them from inadvertent or inappropriate disclosure.

On the PI's surface laptop BitLocker is installed which is a security feature that protects PI's files using data encryption to prevent unauthorized access from hackers and prying eyes. BitLocker provides encryption for full drives and portable drives, and it can protect individual files with data loss protection.

ACCESS RIGHTS
The data will be open access available with the publication; backups will take place on a regular basis and use different storage sites. All the collected data will be inputted and stored in widely available formats (e.g., Excel, Word, PDF, and SPSS) on at least two internal laptops, one OneDrive cloud server and one external hard drive owned by the PI. All of the questionnaire structures will be saved as Word from Qualtrics. To ensure ongoing and long-term security of the data generated during this project, at three copies of all materials will be generated and stored independently from one another at different places (laptops, external hard disc, cloud). All the elaborated publications will be deposited in repositories listed in the Directory of Open Access Repositories (http://www.opendoar.org) with sustainable access at the time of publication. However, if possible, Gold Open Access is preferred and will be the first choice. The data will be used solely for research purposes to be published in peer-reviewed journals; any data reported in such journals will be fully anonymized. All results prepared in the form of publications will be made available via open access. Recognising that ROLLER-COASTER is a funded project, the researcher is deeply committed to advancing sustained open access to scholarly publications and research data, in accord with the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (http://openaccess.mpg.de/Berlin-Declaration). The metadata will be made accessible via open access platforms (e.g., OpenAIRE).

The peer-reviewed publications of the research results will provide all the necessary information to follow and repeat the project. This documentation will be freely available through open access repositories and academic journals to be determined. Open access publications will at least
include details on the applied methodology, reliability and validity analyses, analytical and procedural information, definitions of the chosen items, variables of previous contributions upon which it is built, units of measurement and assumptions made. This type of documentation will help those reviewing the data results to understand and repeat the research project if necessary.

MANAGING SENSITIVE DATA
We will collect age, gender, nationality, educational background, average income of the last three years rounded to thousand USD, experiences, entrepreneurial and other perceptions, along with publicly available venture data related to profit and growth rates. We will send, manage, and store research data using only files and computers, and secure passwords will be created. Passwords will contain at least 10 characters, a mix of upper- and lower-case letters and combinations of numbers and symbols.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

DATA SAVING MANAGEMENT
The data will be stored open access. All the collected data will be inputted and stored in widely available formats (e.g., Excel, Word, PDF and SSPS) on two internal laptops (500 GB each), one OneDrive cloud server (4 TB) and one external hard drive owned by the researcher (4 TB each). All the questionnaire structures will be saved as a word file from Qualtrics. During the research data collection phase, the SPSS data and syntax files will be saved on a regular interim basis. All project documents will be stored at three separate places after the project’s end. As a consequence, the researcher will be responsible for backup and recovery.

What is the long-term preservation plan for the dataset?

BACK UP MANAGEMENT AND STORAGE
To ensure the ongoing and long-term security of the data generated during this project, three copies of all materials will be generated and stored independently at different places (e.g., laptop, external hard disc, cloud, in the ETH library research collection). In general, the main risk related to data security is its loss or destruction, which will be decreased by regular backups. Safe data transfer via email (ETH or Harvard account) will be secured by an Excel file password known to the involved researchers.

Data Sharing

How will you share the data?

DATA ARCHIVE
All the elaborated publications will be deposited in repositories listed in the Directory of Open Access Repositories (http://www.opendoar.org/) with sustainable access at the time of publication. However, Gold Open Access is preferred if possible and will be the first choice. No data will be collected that must be either retained or destroyed. The data will be used solely for research or educational purposes to publish anonymously in peer-reviewed journals; no other usage is planned. Metadata will be stored on three different devices and will be made publicly available (e.g. OpenAIRE). After project closure and when all publications are published a final version of the metadata will be stored at ETH library research collection.

DATA SHARING
During the project is ongoing the research data resulting from this project will not be made available to anyone other than the researchers involved in the project (PI and supervisors). However, all results prepared in the form of publications will be made available via open access. Interesting results will be prepared for publication in top-tier journals, which will be defined throughout the project’s lifecycle. Peer-reviewed research results are made freely available via
the Internet through at least Green Open Access (Gold Open Access will always be preferred). Similarly, open access to the research results will be ensured by self-deposition of accepted manuscripts (after peer review but prior to copy editing and production by the publisher) in any sustainable subject or institutional repository (publisher’s self-archiving policies will be carried out via the Sherpa/Romeo database http://www.sherpa.ac.uk/romeo/), with full respect to the journal’s embargo period.

INFORMING PARTICIPANTS AND THE PUBLIC AT LARGE

After the embargo period of publications or if Gold Open Access is chosen, all publications will be stored on different institutional repositories (e.g., ETHZ). Third parties or other researchers receive access to the full data set after all relevant publications are published. The core project results will be disseminated via local, regional, national, and international offline and online easy-to-read reports and online tools such as a Facebook fan page. All results will be presented anonymously. No data will be able to be connected to specific individuals or institutions.

During ROLLER-COASTER data will be shared as follow:

- Data will be collected via several platforms: Harvard Qualtrics, Prolific and Pavlovia where the PsychoPy online experiment is performed.
- Data will be collected primary by Qualtrics Research Support Service and https://www.prolific.co/.
- We will use the URL shortener bitly.com to implement a quality check
- Harvard Digital Lab for the Social Sciences will also support in data collection.
- The final password protected Excel file and SPSS including all the data will be shared between the PI and supervisors at ETH or Harvard.
- The final Excel file and SPSS including all the data will be made open-access.
- If needed, the Excel file and SPSS file will be shared with research support services at Harvard such as Harvard Statistics Consulting Service | Department of Statistics.
- Results of the data analysis will be shared anonymous in open access publications.
- Data will be stored on PI’s laptop, PI’s onedrive and PI’s external drive.
- The PI will be traveling between Harvard and ETH as allowed due the current COVID situation and in agreement with funding agencies.

Are any restrictions on data sharing required?

RESTRICTION FOR DATA SHARING

The repository or archive in which the publications will be held will be defined at a later. The metadata will be made accessible. A data repository available for free is preferred; as all data sets have long-term value, the data sets will be well-maintained and sustained beyond the lifetime of the project lifecycle. The results of the research project will be shared through peer-reviewed publications in journals and wide-ranging reports for the general public. While the publications will be archived via open access repositories or Gold Open Access methods, the metadata will not be accessible. After project closure and when all publications are published a final version of the meta data will be stored at ETH library research collection and made open access.

Responsibilities and Resources

Who will be responsible for data management?

The PI - Katharina Fellnhofer

What resources will you require to deliver your plan?

Platforms: Pavlovia/PsychoPy and Qualtrics, https://www.prolific.co/ for data collection
Qualtrics Research Support Service