Research Traineeships Proposal 2018-2019

1. Title of the Project: Towards personalized memory diagnostics from large-scale online data

2. Coordinators

Applicant 1: dr. Willem Huijbers, Assistant Professor, Department Cognitive Science and Artificial Intelligence, Tilburg School of Humanities and Digital Sciences, Tilburg, NL. Harvard Medical School, Martinos Center for Biomedical Engineering, Boston, MA, USA.

Applicant 2: dr. Kyle Lang, Assistant Professor, Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, Tilburg, NL

3. Project Summary

A string of failed clinical trials highlights the need to prevent Alzheimer's disease before the stage of cognitive impairment—rather than futilely treating after brain damage is irreversible (Sperling et al. 2014; Scheltens et al. 2016). To prevent Alzheimer's disease, clinicians need better tools for early detection of cognitive decline. However, current cognitive tests fail to distinguish between normal aging and abnormal cognitive decline, largely because tests are not personalized for the individual being tested (Rentz et al. 2013). Individual characteristic influence memory performance—e.g. age, gender, education, language, location, occupation—and this hinders early detection of cognitive decline. Massive online testing now offers the exciting possibility of mapping memory performance to individual characteristics (Germine et al. 2012). Yet the crucial translation from massive online data to a personalized memory examination has not been made.

In the proposed research traineeship, we will conduct pilot work that will help bridge this gap. We will develop an online memory paradigm to measure cognitive performance and individual characteristics in a large online community. The trainees will work on two integrated research projects to develop the online memory paradigm. Project 1 will focus on the design of the online test and the collection of new experimental data, and Project 2 will focus on data analysis—initially using existing datasets collected at Harvard—and the analysis of pilot data. These projects will be tightly integrated; the initial analysis from Project 2 will inform the design choices implemented in Project 1. Together, these projects will demonstrate feasibility and take the first steps towards personalized memory diagnostics based on massive online data.

Methods: An offline version of the proposed memory test was developed by *Applicant 1* for the Harvard Aging Brain study (Huijbers et al. 2016; Dagley et al. 2017). Our famous associative memory exam (FAME) uses images of famous and non-famous individuals to probe recollection of names. Furthermore, *Applicant 1* showed that FAME is sensitive to early biological changes in the brain related to incipient Alzheimer's disease (Huijbers et al. 2016). In parallel, at the psychology department of Harvard University, a group of scientist developed an online platform

(www.testmybrain.org) for online collection of experimental data (Germine et al. 2012). One of the task paradigms used by www.testmybrain.org also employes famous faces, but for a different purpose, namely face recognition (Wilmer et al. 2012). We now aim to bring these two ideas together in a Dutch online version of FAME.

Project 1: Develop an online memory paradigm: In Project 1, we will develop an online famous associative memory examination (e-FAME) tailored to the Dutch population. The primary aim of Project 1 is to develop and evaluate our memory paradigm. This project will be organized into the following steps:

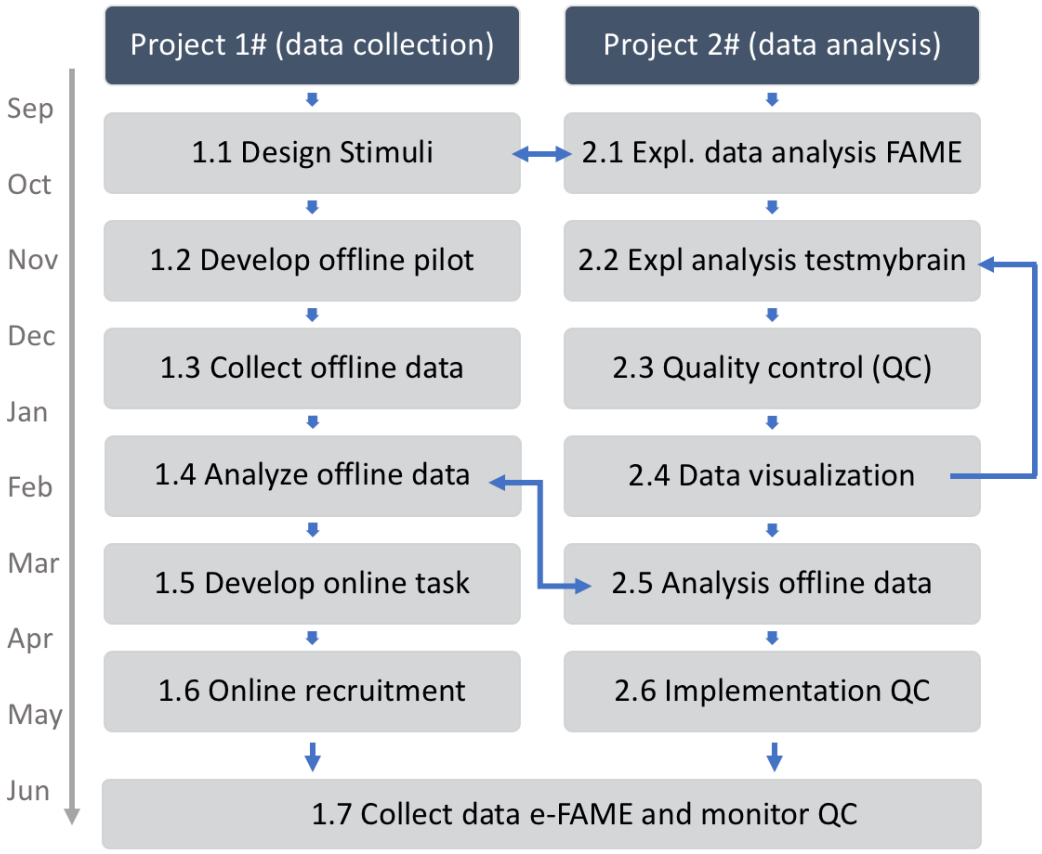
- Step 1.1: Design stimulus materials by gathering images/names of individuals who are famous in the Netherlands (see Project Timeline). In each test, famous items will be randomly sampled from a larger online dataset, so we can later estimate their likelihood of being remembered.
- Step 1.2: Programming the memory test, using software that enables massive online data collection (e.g., the open-source project www.manybrains.org, Qualtrics™).
- Step 1.3: Collect offline data in a traditional lab setting, so we can observe task behavior, detect unforeseen issues, and interview the participants about their experience. We will also query participants for age, gender, education, occupation, language, and personal interest.
- Step 1.4: Analysis of the offline pilot data (see Step 2.6, below).
- Step 1.5: Updating/programming e-FAME. The offline pilot results, together with the input from Project 2, will enable informed design choices.
- Step 1.6: Developing a strategy for online recruitment of participants and designing a basic web portal for data collection.
- Step 1.7: Data collection via e-FAME with continuous, semi-automatic evaluation of data quality (see Step 2.6, below)

Project 2: Data analysis for personalized memory diagnostics: In Project 2, we will evaluate previously collected data in FAME, data from testmybrain.org, and the new pilot data from Project 1. The primary aim of Project 2 is to optimize e-FAME's design by data-driven choices and to integrate online quality control (QC) into the e-FAME system (Step 1.6). This project will be organized into the following steps:

- Step 2.1: Exploratory data analysis on the curated behavioral data from FAME, gathered previously from 120 individuals in Boston (Huijbers et al. 2016). Tasks will include an item-level analysis to identify potential stimulus materials (see Step 1.1).
- Step 2.2: Exploratory data analysis of the dataset from testmybrain.org, gathered from approximately 132.500 individuals online (status on 01/05/2018).
- Step 2.3: Quality control (QC) of the online data, including missing data analyses (Lang & Little, 2018).
- Step 2.4: Data visualization of the results from testmybrain.org. Note, Steps 2.2 - 2.4 are intimately connected and will require multiple iterations before a final evaluation.

- Step 2.5: Analysis of the off-line pilot data from e-FAME (see Step 1.4) and comparison to the results from testmybrain.org.
- Step 2.6: Implementation of tools to monitor the quality of data collection in e-FAME, based on insights gained from the data analyses in Steps 2.2 - 2.5

4. Project Timeline



Deliverables: The first deliverable is high-quality education of the two students who embark on the research traineeship. They will gain hand-ons experience in cutting-edge experimental research, programming, and data analysis techniques. Secondly, we will build the foundation for the infrastructure needed to develop personalized memory diagnostics, including a first version of e-FAME and tools for data visualization and QC. Third, we will collect initial data via e-FAME. The results from the data analyses—of both testmybrain.org and e-FAME—will result in two or more conference abstracts. The comparison of memory data collected in Netherlands and USA will be published as journal article.

Outlook: We will use the data/results to strengthen future grant applications for research on personalized memory diagnostics, methods development, and online data collection. Secondly. we foresee integration of our research within the educational program. Dr. Lang could use the dataset in his DSBG Statistics and Methodology course and dr. Huijbers could employ the online

infrastructure for practical work within Research Workshop. The infrastructure for online data collection can also be used for thesis projects by CSAI students, dataset would be ideal for thesis projects in DSBG and the webportal/task design could be developed further by students in NMD.

5. Research Trainee Profile

Student Profile 1 for traineeship on experimental research (data collection): Requires a motivated student at the bachelor level. We will target a 2nd or 3rd year student from the TSHD program in CSAI. The student will be primarily supervised by dr. Willem Huijbers, who also teaches Research Workshop and Cognitive Neuroscience for CSAI. Requirements: knowledge of programming (html/python) and an affinity for/interest in cognitive science and memory.

Student Profile 2 for traineeship on research analysis (data analysis): Requires a motivated student at the master level, either from the Research Master or Data Science, Business and Governance (to be renamed into Data Science for Society). The student will primarily be supervised by dr. Lang, who also teaches Statistics and Methodology in the DSBG program. Requirements: knowledge of statistics, programming in R or Python and an affinity for/interest in statistical modeling and data mining.

How to apply

Please send your application, motivation letter and CV, to Willem Huijbers w.huijbers@uvt.nl

References

- Dagley A, LaPoint M, Huijbers W, Hedden T, McLaren DG, Chatwal JP, Papp KV, Amariglio RE, Blacker D, Rentz DM, Johnson KA, Sperling RA, Schultz AP. 2017. Harvard Aging Brain Study: Dataset and accessibility. NeuroImage. 144:255–258.
- Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychonomic Bulletin & Review. 19:847–857.
- Huijbers W, Papp KV, LaPoint M, Wigman SE, Dagley A, Hedden T, Rentz DM, Schultz AP, Sperling RA. 2016. Age-Related Increases in Tip-of-the-tongue are Distinct from Decreases in Remembering Names: A Functional MRI Study. Cerebral Cortex.
- Lang KM, Little TD. 2018. Principled Missing Data Treatments. 1–11.
- Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling R, Ferris S. 2013. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. Alzheimers Res Ther. 5:58.
- Scheltens P, Blennow K, Breteler MMB, De Strooper B, Frisoni GB, Salloway S, Van der Flier WM. 2016. Alzheimer's disease. Lancet.
- Sperling R, Mormino E, Johnson K. 2014. The evolution of preclinical Alzheimer's disease: implications for prevention trials. Neuron. 84:608–622.
- Wilmer JB, Germine L, Chabris CF, Chatterjee G, Gerbasi M, Nakayama K. 2012. Capturing specific abilities as a window into human individuality: The example of face recognition. Cognitive Neuropsychology. 29:360–392.