The Role of Context in Human Memory Augmentation

Doctoral Dissertation submitted to the
Faculty of Informatics of the Università della Svizzera italiana
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

presented by

Evangelos Niforatos

under the supervision of

Prof. Marc Langheinrich

April 2018

Dissertation Committee

Prof. Fabio Crestani	Università della Svizzera italiana, Lugano, Switzerland
Prof. Matthias Hauswirth	Università della Svizzera italiana, Lugano, Switzerland
Prof. Enrico Rukzio	Universität Ulm, Ulm, Germany
Prof. Corina Sas	Lancaster University, Lancaster, UK

Dissertation accepted on 20 April 2018

Research Advisor	PhD Program Director
Prof. Marc Langheinrich	**The PhD program Director *pro tempore***

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award.

Evangelos Niforatos
Lugano, 20 April 2018

"*As for me, I am tormented with an everlasting itch for things remote. I love to sail forbidden seas, and land on barbarous coasts.*"

-Herman Melville (1819–1891),
Moby-Dick; or, The Whale (1851).

Abstract

Technology has always had a direct impact on what humans remember. In the era of smartphones and wearable devices, people easily capture on a daily basis information and videos, which can help them remember past experiences and attained knowledge, or simply evoke memories for reminiscing. The increasing use of such ubiquitous devices and technologies produces a sheer volume of pictures and videos that, in combination with additional contextual information, could potentially significantly improve one's ability to recall a past experience and prior knowledge. Calendar entries, application use logs, social media posts, and activity logs comprise only a few examples of such potentially memory-supportive additional information. This work explores how such memory-supportive information can be collected, filtered, and eventually utilized, for generating **memory cues**, fragments of past experience or prior knowledge, purposed for triggering one's memory recall. In this thesis, we showcase how we leverage modern ubiquitous technologies as a vessel for transferring established psychological methods from the lab into the real world, for significantly and measurably augmenting human memory recall in a diverse set of often challenging contexts. We combine experimental evidence garnered from numerous field and lab studies, with knowledge amassed from an extensive literature review, for substantially informing the design and development of future pervasive memory augmentation systems. Ultimately, this work contributes to the fundamental understanding of human memory and how today's modern technologies can be actuated for augmenting it.

Acknowledgements

The author acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET Grant Number 612933.

Contents

List of Figures

List of Tables

Part I

Background

Chapter 1

Introduction

From the very first attempt to preserve a memory by painting it on a cave wall in prehistoric times, to the high-resolution images we capture with our smartphone today, it is evident that we greatly value our memories and that we strive to protect them from fading away over time. Pictures, diaries, notes, and objects of emotional value (i.e., memorabilia) have long since served as reminders of our past. Today's technology has greatly expanded our choices, radically altering the nature and scale of information that we can draw on for remembering. Digital photo albums, progress tracking applications, and e-mail archives are only a few examples of technological artefacts that can improve our "episodic memory" — our ability to remember past experiences, but also our "semantic memory" — our ability to recall attained knowledge.

However, despite this technological proliferation, people still struggle to remember past experiences and previously obtained knowledge. In fact, surprisingly numerous failing memories concern one's daily encounters. How many of us haven't found ourselves admitting: "*I don't remember what I ate yesterday*" or "*I forgot the name of that new guy from the gym*"? This is partially due to today's hectic lifestyle that divides one's daily routine in numerous small segments, while increasingly blurring the line between work and personal life. Inevitably, attention, which is paramount for creating high quality memories, spans over a large number of often simultaneous tasks, and thus hindering the creation of strong memories. The aim of this work is to improve people's ability to remember past experiences and attained knowledge with the use of ubiquitous technologies.

The basic principle for supporting one's memory is to use technology for unobtrusively capturing relevant (contextual) information about one's life, and, subsequently, to use this data to generate and opportunistically present memory-triggering information in the form of so-called memory cues. Memory cues are simply hints (stimuli) that help one recall a past experience or acquired knowledge, such as a picture of a face together with initials for triggering the recollection of the person's full name. Memory cues trigger memory recall and help maximize the details one can remember about a past (personal) event or about something learnt, a process known in Psychology as "*cued recall*". Memory cues are then presented to the user over time in suitable moments such as, on one's smartphone when on the train, or on an ambient display in one's kitchen. Such a presentation of memory cues should improve one's ability to recall a relevant memory over time, an effect that could potentially persist without the need of further technological support (e.g., electronic reminder or online lookup). Given a properly designed review schedule for such cues, one is even able to augment certain memories over others, depending on one's purposes (e.g., upcoming exam).

In this thesis, we focus on visual information (i.e., pictures and videos) for creating effective memory cues, while incorporating any additional contextual information (e.g., location, time, physiological responses, etc.) for the further improvement of visual cues in supporting memory. Particularly, we investigate how contextual information can be captured and used for generating relevant memory cues that enhance people's ability to recall experiences or knowledge. Ultimately, with expertise gained through mobile application prototyping, evidence from subsequent field and lab studies, and design principles from literature review, we contribute to the design and development of future pervasive memory augmentation systems.

1.1 Problem Statement

People forget. A range of reasons for this has been identified in the literature. For example, chronic stress as a result of a hectic lifestyle, for which technology is often greatly responsible, has been found to drastically hinder the formation of strong memories [31, 172]. Depression is another factor that may weaken one's memory, with people who suffer from it being more likely to develop some sort of memory problem [34]. Moreover, one's ability to recall past memories and form new ones is known to progressively deteriorate with ageing in connection with chronic diseases and persistent disorders such as Alzheimer's and Dementia [52]. Lack of exercise has also been linked to memory aggravation, largely attributed to the shrinkage of the *hippocampus* due to reduced blood flow [71] — an area in the brain responsible of the recall of autobiographical memories [228].

Undoubtedly, the negative impact of technology on one's ability to remember cannot be ignored. Either explicitly, through a fast-paced lifestyle, or implicitly, by steadily increasing the technological age-gap that reduces technological benefits for the elders — technology has its share of blame in the formation of fallible memories. However, we believe that technology, via appropriate actuation, holds a significant potential in also improving human memory, and not only the type of memory that holds information about one's past (i.e., episodic), but also the type of memory that stores knowledge (i.e., semantic). We showcase this by adapting modern ubiquitous devices (i.e., smartphones) to support serendipitous "*cued recall*". Cued recall has been found able to identify and aid patients who suffer from memory disorders and diseases such as Alzheimer's and Dementia [107, 232, 250]. However, its use for supporting able-bodied, healthy adults in their everyday life has so far not been systematically explored.

1.2 Research Questions

In our effort to utilize modern technology for aiding remembering, we seek to answer the following four research questions (see Table 1.1).

Section	**Research Question (RQ#)**	**Chapter(s)**
1.2.1	RQ1 – Can event-driven mobile capture produce memory cues that augment episodic memory recall?	4, 5
1.2.2	RQ2 – How does picture capture modality affect both memory recall and memory cue quality?	6
1.2.3	RQ3 – How can the combination of diverse memory cues augment both episodic and semantic memory?	7
1.2.4	RQ4 – Can physiological responses drive memory cue selection for virtual reality memories?	8

Table 1.1. Overview of Research Questions (RQs) and chapters where they are addressed in thesis.

1.2.1 Event-Driven Memory Cue Capture

Perhaps it comes with no surprise that the viewing of memory-supportive content, such as a picture, can trigger one's recollections about a past experience (i.e., episodic recall). Typically, memory cues in the form of pictures are either captured manually (i.e., with a digital camera or a typical smartphone), or at best automatically via lifelogging devices (e.g., SenseCam [111]), and have been particularly successful in assisting patients suffering from memory impairment (e.g., Alzheimer's) [145, 250]. Albeit memory is tightly interconnected with emotion (due to two regions in the brain known as "*amygdala*" and "*hippocampus*" [181]), an emotional memory can neither be stored nor retrieved but instead, can be reconstructed by contextual details residing in episodic memory [189]. Even so, when recalling a (negative) past experience, our recollected emotions are generally positively–skewed, a phenomenon known as the "*rosy view effect*" [161]. It is therefore evident that emotion plays a key role in the recall of autobiographical (episodic) memories. In this work, we introduce and study the notion of "*mobile cued-based augmented memory recall*", where the event-driven memory cue capture, the generation and presentation processes take place entirely on a typical smartphone, for supporting the recollection of episodic memories via elicited emotions. We seek to answer the afore-stated research question (RQ1) by testing our (mobile) memory interventions outside a typical Psychology lab, in our participants' everyday life settings. In particular, we present memory cues on our participant's smartphones at opportune moments throughout the day (i.e., when at the bus stop), and systematically evaluate their ability later to recall a past experience. At this stage, our aim is to evaluate the feasibility of our approach as a proof of concept.

Next, we inquire into how can memory cues, produced via event-driven mobile capture, augment episodic memory in practice, in diverse contexts, and with manifold purposes. An interesting venue to explore is the potential of cue-based augmented memory recall to facilitate the process of User Experience (UX) evaluation. In fact, it is broadly believed that emotion is a key aspect of UX, in the field of Human-Computer Interaction (HCI) and Ergonomics, since emotion plays a critical role in the design of better systems [103, 104, 157]. As a result, a plethora of methods has been proposed for measuring emotion in UX, with the Experience Sampling Method (ESM) considered as the "*gold standard*" for in situ measurement [142]. However, ESM does entail significant drawbacks (e.g., highly obtrusive), that led experts to introduce the Day Reconstruction Method (DRM) [124]. The DRM relies on one's memory for retrospectively eliciting exhibited emotions, feelings, thoughts, etc., and thus it has been proposed as a viable alternative to ESM for measuring UX. Nevertheless, the DRM hinges solely on one's organic memory without any external support, simply by asking one to recall past experiences in a chronological order and in the form of episodes. Therefore, an opportunity emerges for utilizing our "*mobile cue-based augmented memory recall*" approach in the elicitation of emotions for assisting the process of UX evaluation on mobile and automotive platforms.

In this thesis, we first assess the concept of mobile cue-based augmented memory recall in the wild, with memory cues produced by event-driven capture, and then showcase its methodological potential as a retrospective UX evaluation method. In fact, the notion of retrospective UX evaluation may be particularly adequate for highly diverse and challenging contexts, where established methods (e.g., ESM) are considered disruptive (i.e., mobile UX evaluation) or even perilous (i.e., automotive UX evaluation). All in all, we expect that the memory improvement effect, introduced by our memory interventions, will persist over time, even in the absence of further memory support. The vision of a nearly "time-proof" memory provides ample ground for exploring memory recall augmentation in proliferating application scenarios and contexts, eventually informing the design and implementation of future pervasive memory augmentation systems.

1.2.2 Picture Capture Modality Effect

In this work, picture capture is considered integral for producing effective memory cues. In fact, picture capture, typically manual but lately also automatic, has long been the primary way we preserve our memories. However, recent findings show that the very act of manually capturing a picture can be detrimental to memory formation (i.e., Henkel's "*photo–taking impairment effect*" [108, 211]). Automatic picture capture, with the use of lifelogging cameras, could potentially alleviate this effect, enabling one to fully immerse oneself into the experience. Nevertheless, automatic picture capture produces a sheer amount of content that is cumbersome for one to review, and hence experts have been stressing the "*need for selectivity, not total capture*" [210]. To this end, no prior study has thoroughly investigated the implications of manual and automatic picture capture modalities on human memory as well as, the quality of the generated memory cues via these modalities. For example, the source of Henkel's "*photo–taking impairment effect*" has not been thoroughly pinpointed yet. In this work, we explore whether the "*photo–taking impairment effect*" manifests due to disruption at encoding by the act of manual picture capture [108, 211], or perhaps by the possibility that the camera is treated as an external memory support (e.g., "the Google effect" [212]).

For trialling the notion of "*selectivity, not total capture*" in the real world, we developed a custom mobile picture capture application that enabled us to investigate the plausible effects of an imposed capture limitation on human memory and the quality of the generated memory cues. Our intention here is to explore new design dimensions that will inform the design and development of future pervasive memory augmentation systems.

1.2.3 Hybrid Episodic–Semantic Memory Cues

After evaluating the concept of automatic and event-driven capture in producing memory cues that support the recollection of autobiographical (episodic) memories, we quickly realized that it could be useful in a wider set of everyday life scenarios and contexts. As such, we investigated the potential of cue-based augmented memory recall to improve efficiency and overall experience in the workplace. Quite often, the capture of contextual information can result in memory cues that trigger both the episodic memory (memory about self), such as pictures, and the semantic memory (memory about facts), such as text, hence allowing us to distinguish between episodic and semantic memory cues for supporting hybrid episodic–semantic recollections [227]. In fact, episodic and semantic memory components operate so tightly in tandem, that when either an episodic or semantic memory is recalled, both actions are registered as a new episodic memory [227]. Such hybrid recollections are particularly prevalent in contexts where encountering novel experiences is intertwined with acquiring new knowledge (e.g., classrooms, academic settings, modern workplace, etc.). Among others, work meetings pose an ideal opportunity for trialling the combinatory power of episodic and semantic memory cues to synergistically support memory recall, especially when no prior quantifiable baseline has been established.

In this thesis, we highlight the utilitarian facet of cued–based augmented memory recall by utilizing hybrid memory cues that trigger both episodic and semantic memory recall, for supporting memory recall in the workplace. To the best of our knowledge, no prior work has systematically delivered and quantified memory augmentation in work meetings. Therefore, the achieved memory augmentation baseline, reported in this work, can be used as a reference, against which future pervasive memory augmentation systems could compare.

1.2.4 Physiological Responses in VR Memory Cue Selection

As we have seen in the previous sections, automatic picture capture, typically implemented by lifelogging cameras (e.g., SenseCam [111]), is a great source of memory cues, but it comes at the cost of storing, searching and processing voluminous data. Prior work has demonstrated the use of physiological responses, such as Electro–dermal Activity (EDA), for selecting the most personally relevant pictures throughout one's daily life and hence, supporting richer recollections (e.g., AffectCam [200]). The proliferation of wearable devices (e.g., Apple Watch, Empatica E4, etc.) has increased the spectrum of physiological responses (e.g., heart rate variability, skin temperature, blood volume pulse, etc.) one could utilize for improving the memory cue selection process, yet their potential in that respect still remains untapped.

In this thesis, we investigate the interplay of physiological responses with memory recall in the simulated and controlled settings of Virtual Reality (VR). A novel and immersive VR experience provides an excellent test bed for acquiring feedback about the potential of physiological responses to serve as "memory biomarkers", indicators of strong or weak memory formation for driving the selection of memory cues. In fact, any insights gained can be utilized for transferring methods and prototypes from the lab into the real world, with the ultimate aim of designing and developing better future pervasive memory augmentation systems.

1.3 Thesis Goals

The goal of this thesis is to contribute to the design and development of (mobile) applications for enhancing one's ability to recall a past experience and prior knowledge. This is done by drawing on experimental evidence stemming from a series of studies conducted with mobile and desktop application prototypes, both in lab and in everyday life settings. Among others, our contributions consist of capture techniques for mobile platforms, content filtering strategies, and design guidelines that aim at benefiting the design and development of future pervasive memory augmentation systems. In overall, the effectiveness of our approach (i.e., cue-based augmented memory recall) is demonstrated by testing prototypes in a largely diverse set of contexts encountered in everyday life such as, on the go, in the car, at the workplace, at the campus, and others. A summary of contributions of this thesis is shown in Table 1.2 below.

Nr.	Contribution	Chapter(s)	Publication(s)
1	Design principles for Lifelogging User Interfaces (LUIs)	2	—
2	Mobile cued–based memory augmentation	4	[165]
3	A retrospective UX evaluation method	5	[165, 166, 251]
4	Impact of picture capture on human memory	6	[164, 167]
5	A quantifiable baseline for memory augmentation in work meetings	7	[169]
6	Hybrid episodic–semantic memory cues	7	[169]
7	Physiological responses for driving memory cue selection in VR experience recall	8	[153, 154, 168]

Table 1.2. Overview of contributions of this thesis.

1.4 Research Context: The RECALL Project

The principal component of the work described in this thesis was conducted during the time the author worked in Ubiquitous Computing (UC) group[1] at Università della Svizzera italiana (USI) located in Lugano, Switzerland, under the supervision of Prof. Marc Langheinrich, and in the frames of the RECALL project. The RECALL[2] EU project received 3-year funding via the Future Emerging Technologies (FET) programme and the 7th Framework Programme for Research of the European Commission, under the FET grant number 612933. The RECALL project consortium was comprised of 4 partner universities (i.e., USI, Lancaster University, University of Stuttgart, and Essex University), that closely collaborated for "*re-thinking and re-defining the notion of human memory augmentation*" with the use of ubiquitous technologies. Particularly, the partners worked together over the course of 3 years (i.e., November 2013–October 2016) for prototyping and trialling technological interventions that enhance the memory of healthy individuals in everyday scenarios and settings. Among others, the collaboration resulted in conjoint publications such as [60] and at the Ubicomp 2017 conference [164]. Moreover, we co-organized a series of workshops such as workshop for augmenting human mind (WAHM [61]) and workshop on mobile cognition (WMC [59]), hosted at Ubicomp 2016 and MobileHCI 2015 conferences, respectively. All work reported in this thesis is the product of conjoint effort with colleagues (RECALL-related and non-related). Work that the author has led (i.e., as the primary author) is described in detail, whereas co-authored work is briefly summarized (for a full list of publications see Table 1.3). Author's as well as collaborators' contributions are outlined in the beginning of all chapters that entailed user studies.

1.5 Ethics

The experimental work conducted in the frames of this thesis entailed significant ethical implications, as it tested interventions with (healthy) *human subjects*, it investigated the *effects of memory augmentations* on participants' ability to recall

[1] `https://uc.inf.usi.ch` *(all URLs current as of May 14, 2018).*
[2] `http://recall-fet.eu`

a past experience or prior knowledge, and it involved the *collection of personal and identifiable data*. As such, our experimental protocols and study designs have been reviewed by an assigned ethical board, assessing the adequateness of the employed methods and techniques before and for each experiment and/or study we have conducted in this work. In the following sections, we briefly describe how we adhered to established sound ethical practices [140].

1.5.1 Human Subject Experiments

Research on human subjects should follow established research ethics principles for physically and mentally protecting participants in research experiments, as stated for the first time the Nuremberg Code[3] (1947), subsequently in the Declaration of Helsinki (1964, revised in 1975, in 1996, and in 2000), and in the Belmont Report[4] (1974) [4]. Clearly, the initial research ethics principles have been amended over the years for keeping up with scientific and societal progress. Apart from ensuring at all times the safety of participants, perhaps the most notable ethical requirements are those of informed consent and devotion to research that benefits society.

The research work described in this thesis involved experiments on healthy human subjects with the main objective to augment their (episodic and semantic) memory via technology. These included both lab experiments, in-situ and in-the-wild experiments, and they concerned both behavioural testing and usability testing. Behavioural testing, in the context of this endeavour, included experimental work on the ability of subjects to recall past experiences and prior knowledge. These experiments contrasted the amplified recall capabilities of experimental participants, who used the proposed interventions, with control participants who did not. Not only have experimental tasks and experimental stimuli been employed in recall tests, but participants have been asked to fill in self-report questionnaires that surveyed their habits, preferences and evaluation of the employed memory intervention, as a recall augmentation to their own recall capacities. Testing involved the use of technologies we use daily (i.e., smartphones and smart watches), plus non-invasive physiological sensing devices, such as eye-trackers, and physiological responses measuring wristbands.

Usability tests validated the usability and utility of the employed experimental interventions. Participants have been asked to perform a set of tasks and fill in self-report questionnaires, or answer interview questions. Task performance

[3] https://history.nih.gov/research/downloads/nuremberg.pdf

[4] https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

(e.g., recall scores, completion times and rates), and experience sampling inputs (e.g., emotion self-assessment) have been recorded for further analysis. In all human subject experiments conducted in the frames of this work, the utmost care has been taken in maintaining personal integrity (i.e., appropriate use of equipment, information, consent, benefits and risks, anonymity, options for opting out and data, protection), and safety.

1.5.2 The Effect of Memory Experiments

It is probable that the memory experiments conducted in the frames of this work may at times have had no effect, or even a temporary detrimental effect on participants' memory recall capacity, instead of an anticipated memory recall augmentation. For example, a popular user study scenario in our applied experimental methodology required participants to perform a recall task, after their recall capacity had been amplified by being exposed to an experimental memory intervention. In the first stages of this work, our memory interventions were not particularly successful, due to selecting suboptimal or irrelevant memory cues for augmenting one's memory recall. As a result, participants received little, if any, memory augmentation benefit, and when repeatedly asked to recall the specifics of the a past recollection undergoing a so-called augmentation, may have felt distress or fatigue. In the subsequent stages of this work, our interventions became gradually more effective in assisting one's recall of a past experience and prior knowledge. In turn, we expect that this has significantly affected our participants' recall ability in amplifying specific memories in the innate cost of attenuating other related ones (see "retrieval-induced forgetting effect" [1]). However, prior work has shown that forgetting is a rather natural and necessary process during the time memory consolidation processes take place [201]. Since most of our experiments took place outside a typical Psychology lab and involved participants in their daily whereabouts, we were thus unable to quantify forgetting of other related memories as a collateral effect of our targeted memory augmentation interventions. We expect though that "*collateral forgetting*" could be reversed simply by re-targeting those attenuated memories and applying the same memory augmentation approach that this work introduces[5]. To this end, it could be argued that a memory augmentation intervention may in fact have been too coarse in the first stages of this work, or temporarily causing the attenuation of related but non-targeted memories, but under no circumstances could be potentially "harmful".

[5] Given that memory cues are available for augmenting attenuated memories in the first place.

1.5.3 Personal and Identifiable Data

In this research work, we have collected personal information and identifiable data in different stages, as part of experiments and pilot prototype deployments. Data has been collected implicitly from system log files, automatically via automatic capture equipment (e.g., a lifelogging camera) as well as, explicitly using physiological measuring equipment, questionnaires, interviews, field notes, and audio and video recordings. Developed interventions and prototypes have also collected data for fulfilling their task (e.g., keeping record of a participant's daily activities, cognitive and physiological states).

The interactions of subjects with a prototype as well as, their demographic background information have been collected in order to evaluate an intervention's memory augmentation performance and its usability. The identity of the subject, typically contact information such as name and e-mail addresses, have been stored separately from data collected during individual user studies. Subjects have been known within and across studies only by a pseudonymous identifier. In addition, we have also collected various background demographics in order to better classify and understand experimental results. This data is not personally identifiable. The exact demographic information collected had been determined at the time of experimental design, but typically included age, gender, education, and occupation. The exact data collected has always been detailed in the ethical worksheet drafted before and for each experiment. All data collections have been conducted according to the corresponding national and institutional regulations, including notifying the appropriate national institutional organizations when required, and always after the approval of the assigned ethical board.

The personal information of study subjects (i.e., their identity information) have been stored in a centralized fashion and in encrypted form. Any user profile created during the interaction with a memory augmentation intervention has been encrypted within the system, with access keys only available to subjects and the involved researchers. Throughout the duration of this work, standard encryption techniques have been used for storing data collected locally or remotely during experiments.

1.6 Thesis Outline

This thesis is comprised of 12 chapters, organized in 3 parts, followed by a bibliography section. In Part I – Background, we present the motivation, theoretical underpinning and methodological background of this work. Part II – Studies, encompasses the research conducted for augmenting episodic and semantic memory recall. In the last part (Part III – Synthesis), we draw on obtained knowledge from field deployments and literature review for answering our research questions, we describe our contributions, and we summarize the outcome and the future of this work. Below, we provide a brief description for each chapter in each part of this thesis:

1.6.1 Part I – Background

- **Chapter 1 – Introduction.** The first chapter describes our vision and motivation for augmenting memory recall with ubiquitous technologies, defines the research questions this thesis answers, outlines the research context based on which this thesis is written, and summarizes the key challenges we faced throughout this endeavour, including overall contribution.
- **Chapter 2 – Foundations.** In the second chapter, we provide a synopsis on the structure of human memory and the processes that pertain to it. Then, we showcase how modern (technological) trends and practices such as lifelogging and the quantified-self movement could support memory recall through the appropriate actuation of captured contextual information. We also include insights gained from an extensive literature review on Lifelogging User Interfaces (LUIs), eliciting design principles and guidelines to which LUIs adhere and the goals they fulfil.
- **Chapter 3 – Augmenting and Measuring Memory Recall.** In this chapter we describe our methodology and in particular how we transferred an established psychological method such as "cued recall", from the lab into the wild. We present the apparatus we used, and we outline the methods we applied for collecting contextual information, generating memory cues, delivering memory cues, and finally evaluating the effectiveness of our memory-augmenting interventions.

1.6.2 Part II – Studies

- **Chapter 4 – Event-Driven Image Capture for Augmenting Episodic Memory Recall.** Here we describe a first user study that investigated the potential of event-driven captured self-face pictures (event-driven selfies), taken at a common mobile interaction moment (i.e., screen unlock) via a dedicated mobile application that utilizes the front–facing camera of a typical smartphone, to effectively trigger episodic memory recall. The potential of event-driven selfies in supporting episodic memory recall was tested in several stages after capture.

- **Chapter 5 – Augmenting Memory Recall for UX Evaluation.** In this chapter, we demonstrate how event-driven captured selfies in the form of memory cues can benefit the field of retrospective UX evaluation for mobile applications. We present a user study with regular commuters predicting, experiencing and recalling negative emotions (i.e., anger and frustration) during the experience of daily commute. Particularly, we were able to unveil plausible effects on how commuters recall a negative driving experience with potential significant implications for in-car UX design. Leveraging on the the potential of selfies to trigger memory recall, we showcase an application prototype that features a combination of memory cues (i.e., pictures, video of facial expressions, location, etc.) for safely eliciting driver's affective state at a later stage. We believe our prototype can benefit urban planners and in-car UX designers by retrospectively measuring in-car UX levels.

- **Chapter 6 – Capture Modality Effect on Memory Recall.** Here, we present an extensive user study that compared three different picture capture modalities (i.e., unlimited, limited, and automatic) in their effect on memory recall, with and without the support of the captured pictures as memory cues. The study explored novel design dimensions for future pervasive memory augmentation systems, by artificially limiting the amount of pictures one can take via a dedicated mobile application, while comparing it with prevalent mobile picture capture, and a popular wearable lifelogging camera.

- **Chapter 7 – Augmenting Memory Recall for Work Meetings.** In this chapter, we report a user study that inquired into the effectiveness of recall-augmenting interventions in the workplace context by recruiting participants that meet regularly in a "supervisor-supervisee" fashion. By utilizing a combination of episodic and semantic memory cues, we were able to establish a quantifiable baseline of memory augmentation in work meetings that serves as reference for future research on pervasive memory augmentation systems.

- **Chapter 8 – Physiological Responses in VR Experience Recall.** Here, we present two studies that highlight the potential of physiological responses in driving the selection of memory cues that could trigger episodic memories at later stages. We investigate the interplay of physiological responses with the recall of past VR experiences by recruiting participants in mixed educational and recreational contexts (i.e., a museum and an aquarium). We also present another wearable prototype that could shed light on the role of physiological responses in memory recall in the real world.

1.6.3 Part III – Synthesis

- **Chapter 9 – Contribution Summary.** In this chapter, we combine the knowledge garnered throughout this thesis for answering the research questions raised in Chapter 1 – Introduction, while providing guidelines for the design and development of future pervasive memory augmentation systems. We also present here any limitations concerning our approach.

- **Chapter 10 – Conclusion and Future Work.** In the last chapter, we reflect on the implications that derive from the use of memory augmentation technologies in everyday life settings. We conclude by sharing our vision on a future software architecture that aims at amplifying not only our memory, but our cognitive capacities in a holistic fashion.

1.7 Publication Overview

The work conducted in the context of this thesis has been published in peer-reviewed conferences, journals and magazines. For an overview, see Table 1.3 below.

Nr.	Publications
1.	Karapanos, E., Barreto, M., Nisi, V., & **Niforatos, E.** [2012]. Does locality make a difference? Assessing the effectiveness of location-aware narratives. *Interacting with Computers, 24*(4), 273-279.
2.	Gouveia, R., **Niforatos, E.**, & Karapanos, E. [2012]. Footprint Tracker: reviewing lifelogs and reconstructing daily experiences. arXiv preprint arXiv:1207.1818.
3.	Alves, R., Lim, V., **Niforatos, E.**, Chen, M., Karapanos, E., & Nunes, N. J. [2012]. Augmenting Customer Journey Maps with quantitative empirical data: a case on EEG and eye tracking. arXiv preprint arXiv:1209.3155.
4.	**Niforatos, E.**, Karapanos, E., Alves, R., Correia Martins, M. C., Chen, M., & Nunes, N. [2013]. Enwildering the lab: merging field evaluation with in-lab experience sampling. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 313-318). ACM.

5.	**Niforatos, E.**, Langheinrich, M., & Bexheti, A. [2014]. My good old kodak: understanding the impact of having only 24 pictures to take. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1355-1360). ACM.
6.	**Niforatos, E.**, & Karapanos, E. [2015]. EmoSnaps: a mobile application for emotion recall from facial expressions. *Personal and Ubiquitous Computing, 19*(2), 425-444.
7.	Wurhofer, D., Krischkowsky, A., Obrist, M., Karapanos, **E., Niforatos**, E., & Tscheligi, M. [2015]. Everyday commuting: prediction, actual experience and recall of anger and frustration in the car. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 233-240). ACM.
8.	**Niforatos, E.**, Karapanos, E., Langheinrich, M., Wurhofer, D., Krischkowsky, A., Obrist, M., & Tscheligi, M. [2015]. eMotion: retrospective in-car user experience evaluation. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 118-123). ACM.
9.	Bexheti, A., Fedosov, A., Findahl, J., Langheinrich, M., & **Niforatos, E.** [2015]. Re-Live the Moment: Visualizing run experiences to motivate future exercises. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 986-993). ACM.
10.	**Niforatos, E.**, Lim, V., Vuerich, C., Langheinrich, M., & Bexheti, A. [2015]. PulseCam: Biophysically Driven Life Logging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 1002-1009). ACM.
11.	Dingler, T., Bexheti, A., **Niforatos, E.**, & Alt, F. [2015]. Workshop on Mobile Cognition: Using Mobile Devices to Enhance Human Cognition. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 970-973). ACM.
12.	Dingler, T., El Agroudy, P., Le, H. V., Schmidt, A., **Niforatos, E.**, Bexheti, A., & Langheinrich, M. [2016]. Multimedia Memory Cues for Augmenting Human Memory. *IEEE MultiMedia, 23*(2), 4-11.
13.	Bexheti, A., **Niforatos, E.**, Bahrainian, S. A., Langheinrich, M., & Crestani, F. [2016]. Measuring the effect of cued recall on work meetings. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (pp. 1020-1026). ACM.

14.	Dingler, T., Kunze, K., **Niforatos, E.**, Gurrin, C., Giannopolos, I., Dengel, A., & Kise, K. [2016]. WAHM 2016: 3rd workshop on ubiquitous technologies for augmenting the human mind. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 1010-1013). ACM.
15.	Fedosov, A., **Niforatos, E.**, Elhart, I., Schneider, T., Anisimov, D., & Langheinrich, M. [2016]. Design and evaluation of a wearable AR system for sharing personalized content on ski resort maps. In Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (pp. 141-152). ACM.
16.	Marchiori, E., **Niforatos, E.**, & Preto, L. [2017]. Measuring the Media Effects of a Tourism-Related Virtual Reality Experience Using Biophysical Data. In *Information and Communication Technologies in Tourism 2017* (pp. 203-215). Springer, Cham.
17.	**Niforatos, E.**, Cinel, C., Mack, C. C., Langheinrich, M., & Ward, G. [2017]. Can Less be More?: Contrasting Limited, Unlimited, and Automatic Picture Capture for Augmenting Memory Recall. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(2), 21.
18.	**Niforatos, E.**, Fedosov, A., Elhart, I., & Langheinrich, M. [2017]. Augmenting Skiers' Peripheral Perception. In Proceedings of the 2017 ACM International Symposium on Wearable Computers (pp. 114-121). ACM.
19.	**Niforatos, E.**, Vourvopoulos, A., & Langheinrich, M. [2017]. Amplifying human cognition: bridging the cognitive gap between human and machine. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 673-680). ACM.
20.	Vourvopoulos, A., **Niforatos, E.**, Hlinka, M., Škola, F., & Liarokapis, F. [2017]. Investigating the effect of user profile during training for BCI-based games. In *Virtual Worlds and Games for Serious Applications (VS-Games), 2017* (pp. 117-124). IEEE.
21.	**Niforatos, E.**, Laporte, M., Bexheti, A., & Langheinrich, M. [2018]. Augmenting Memory Recall in Work Meetings: Establishing a Quantifiable Baseline. In *Proceedings of the 9th Augmented Human International Conference*, p. 4. ACM, 2018.
22.	Marchiori, E., **Niforatos, E.**, & Preto, L. [2018]. Analysis of users' heart rate data and self-reported perceptions to understand effective virtual reality characteristics. *Information Technology & Tourism*, 1-23.

Table 1.3. Relevant work published so far.

1.8 Additional Publications

The table below (Table 1.4) lists additional publications that were completed during this thesis. These publications are either a product of M.Sc. and B.Sc. student supervision or originate from other collaborations and projects, such as Atmos[6] and MYGEOSS[7].

Nr.	Publications
1.	**Niforatos, E.**, Karapanos, E., & Sioutas, S. [2012]. PLBSD: a platform for proactive location-based service discovery. *Journal of Location based services, 6*(4), 234-249.
2.	**Niforatos, E.**, Vourvopoulos, A., Langheinrich, M., Campos, P., & Doria, A. [2014]. Atmos: a hybrid crowdsourcing approach to weather estimation. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 135-138). ACM.
3.	Elhart, I., Scacchi, F., **Niforatos, E.**, & Langheinrich, M. [2015]. ShadowTouch: A Multi-user Application Selection Interface for Interactive Public Displays. In *Proceedings of the 4th International Symposium on Pervasive Displays* (pp. 209-216). ACM.
4.	**Niforatos, E.**, Fouad, A., Elhart, I., & Langheinrich, M. [2015]. WeatherUSI: crowdsourcing weather experience on public displays. In *Proceedings of the 4th International Symposium on Pervasive Displays* (pp. 241-242). ACM.
5.	**Niforatos**, E., Elhart, I., & Langheinrich, M. [2015]. Public displays for monitoring and improving community wellbeing. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 775-778). ACM.
6.	Fedosov, A., **Niforatos, E.**, Alt, F., & Elhart, I. [2015]. Supporting interactivity on a ski lift. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 767-770). ACM.
7.	**Niforatos, E.**, Vourvopoulos, A., & Langheinrich, M. [2015]. Weather with you: evaluating report reliability in weather crowdsourcing. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia* (pp. 152-162). ACM.

[6] `http://myweather.mobi`

[7] `http://digitalearthlab.jrc.ec.europa.eu/activities/mygeoss-applications-your-environment/57752`

8.	Fedosov, A., Elhart, I., **Niforatos, E.**, North, A., & Langheinrich, M. [2016]. SkiAR: Wearable augmented reality system for sharing personalized content on ski resort maps. In *Proceedings of the 7th Augmented Human International Conference 2016* (p. 46). ACM.
9.	**Niforatos, E.**, Elhart, I., Fedosov, A., & Langheinrich, M. [2016]. s-Helmet: A Ski Helmet for Augmenting Peripheral Perception. In *Proceedings of the 7th Augmented Human International Conference 2016* (p. 45). ACM.
10.	**Niforatos, E.**, Elhart, I., & Langheinrich, M. [2016]. WeatherUSI: User-Based Weather Crowdsourcing on Public Displays. In *International Conference on Web Engineering* (pp. 567-570). Springer International Publishing.
11.	Fedosov, A., Ojala, J., **Niforatos, E.**, Olsson, T., & Langheinrich, M. [2016]. Mobile first?: understanding device usage practices in novel content sharing services. In *Proceedings of the 20th International Academic Mindtrek Conference* (pp. 198-207). ACM.
12.	**Niforatos, E.**, Vourvopoulos, A., & Langheinrich, M. [2017]. Understanding the potential of human–machine crowdsourcing for weather data. *International Journal of Human-Computer Studies, 102*, 54-68.
13.	Landoni, M., Fedosov, A., & **Niforatos, E.** [2017]. Promoting CARE: Changes via Awareness, Recognition and Experience. In *Proceedings of the 2017 Conference on Interaction Design and Children* (pp. 783-787). ACM.
14.	**Niforatos, E.**, Fedosov, A., Langheinrich, M., & Elhart, I. [2018]. Augmenting Humans on the Slope: Two Electronic Devices That Enhance Safety and Decision Making. *IEEE Consumer Electronics Magazine*, 7(3), 81-89.

Table 1.4. Additional work published so far.

Chapter 2

Foundations

Evolution has always been the main driving force of change for both the human body and brain. Presently, in the Information era, our cognitive capacities (attention, memory, learning, etc.) cannot simply rely on natural evolution to keep up with the immense advancements in the field of ubiquitous technologies, which remain largely uninformed about our cognitive states. As a result, a so-called "*cognitive gap*" is forming between the human (user) and the machine (system), preventing us from fully harnessing the benefits of modern technologies [170]. In this thesis, we focus on how we can bridge this cognitive gap by utilizing ubiquitous capture and presentation technologies for enhancing one critical aspect of our cognition: *human memory*. In fact, ongoing research and development of near–constant capture devices, unlimited storage, and algorithms for retrieval, and the resulting personal data, have opened the door to a vast range of applications that could significantly assist memory recall in everyday-life settings [60, 111]. In the following sections, we first describe how human memory is structured and how it works, with particular emphasis on the two major human memory components on which this work draws: **episodic** and **semantic** memory. Next, we discuss how **lifelogging**[1] as a practice can augment memory recall through appropriate actuation, and we elicit design principles for Lifelogging User Interfaces (LUIs). Finally, we go beyond lifelogging to present our vision for utilizing capture technologies for holistic **human mind amplification**.

[1]Part of this section includes input from the M.Sc. thesis titled "*A Classification Scheme of Lifelogging User Interfaces*" by Daniel Cardozo, whom the author co-supervised with Prof. Marc Langheinrich and Prof. Fabio Crestani.

2.1 Human Memory

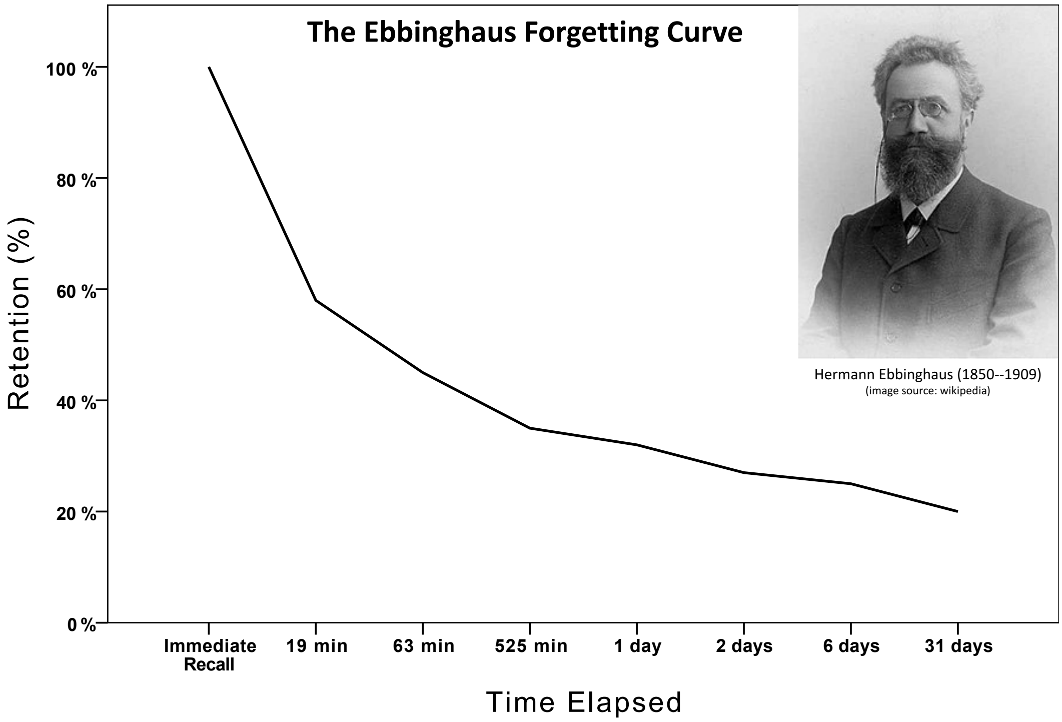


Figure 2.1. The Ebbinghaus forgetting curve describes the decline of memory retention over time. In particular, the curve showcases how information is lost over time when there is no attempt to retain it (e.g., through a rehearsal).

Human Memory has been methodically investigated since 1880s, when the German philosopher Herman Ebbinghaus introduced the first scientific approach to studying memory and attempted to model how information retention deteriorates over time [11] (see Figure 2.1). Ever since, human memory has been the subject of scientific research and debate in multiple fields such as Psychology, Cognitive Psychology, and Neuroscience, with each one providing a different perspective and thus, different understanding on what human memory is and how it operates. Consequently, there are many approaches and models that attempt to describe how human memory is structured, how it stores and retrieves information, in which parts of the human brain it is located, and other aspects. As the basis of this work, we consider a rather simple but very widespread human memory model: the Atkinson and Schiffrin's *Multi-store Model of Human Memory* [6] (Figure 2.2). The multi-store model theorizes the existence of 3 main memory components, namely *sensory memory*, *short-term memory* (also called *working* *memory*) and *long-term memory*. The long-term memory is further divided into *explicit* and *implicit* memory components. In the explicit memory branch, under the further classification as declarative memory, reside the **episodic memory** and **semantic memory** components, on which this work draws. In the following sections, we briefly describe the multi-store model, the different memory components of which it is comprised, and the memory processes that pertain to it, primarily from a Psychology perspective, with sporadic references to the field of Neuroscience.

2.1.1 The Multi-Store Model

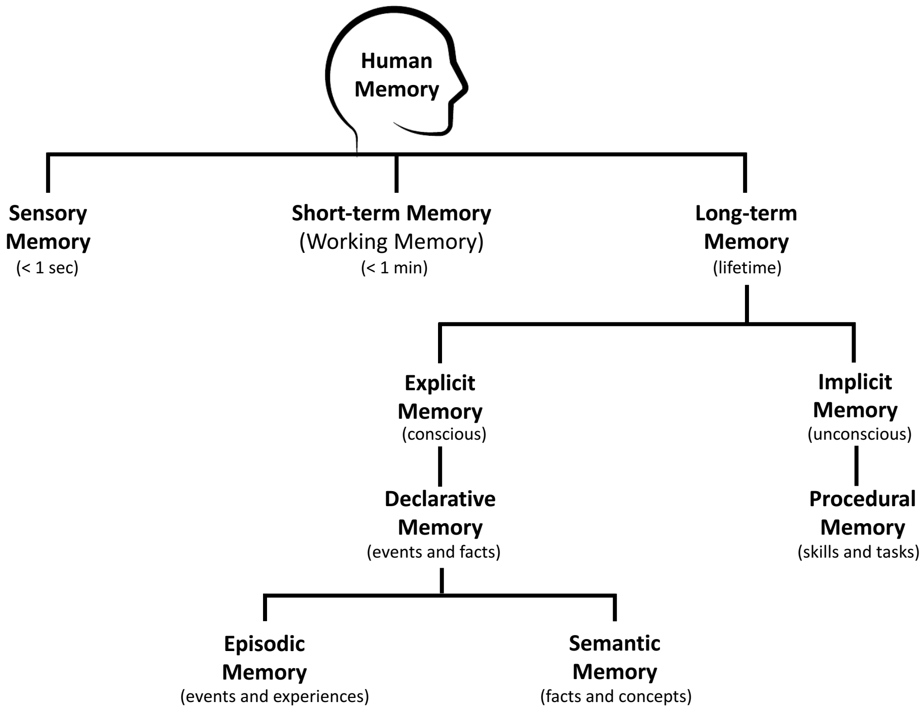


Figure 2.2. The structure of human memory as a sequence of three stages (i.e., Sensory → Short-term → Long-term) rather than a unitary process, known as the multi-store or Atkinson-Shiffrin model, after Richard Atkinson and Richard Shiffrin who introduced it in 1968 [7].

On one hand, the multi-store model is heavily influential receiving support by a multitude of studies in the distinction between short-term and long-term memory and their characteristics (i.e., encoding, duration and capacity), while it can account for primacy and recency effects [162]. On the other hand, the model is considered oversimplified, passive, one-way or linear model, mainly in that it suggests that both long-term and short-term memory each operate in a single, uniform fashion. Among others, criticism focuses on working memory being more than just a simple unitary component, and instead comprised of different subcomponents (e.g., central executive, visual–spatial, etc.) [12]. Particularly for long-term memory, critics proclaim that it is rather unlikely that different types of stored information such as remembering one's online banking password and remembering what one ate yesterday end up being stored within a unitary/single memory buffer/store. Albeit, the model makes a distinction between semantic memory (i.e., facts), episodic memory (i.e., experiences) and procedural memory (i.e., skills) components (see Figure 2.2). Furthermore, the multi-store model has also been criticized for overestimating the importance of attention and rehearsing during the memory formation process, and neglecting additional process elements. Interestingly though, the structure of the multi-store model comprising sets of stores/buffers with information flowing through them, same as information flows through a system, resembles greatly an information processing model with an input, process, and output components (i.e., von Neumann model/architecture). Craik et al., provide a comprehensive overview on the advantages and disadvantages of the multi-store memory model [53]. Despite the aforementioned flaws, we decided to base our work on the multi-store model, since it offers an intuitive, simple and interdisciplinary–compatible systemic view over human memory, without the need to greatly deepen our knowledge on the specificities of Human Memory theory, Cognitive Psychology and Neuroscience. According to the multi-store model, there are three main memory types/components, *sensory memory*, *short-term memory* and *long-term memory*, as described briefly next.

Sensory Memory. It is considered the shortest-term component of human memory and has the ability to retain impressions of sensory information after the original stimuli have ended for a duration of 0.25–0.5 seconds [188]. It operates as a store/buffer for storing very large volume of information caused by stimuli perceived by our five senses which are retained extremely accurate but very briefly. As such, the ability to look at something for a split second and remember how it looked like (e.g., shape, colour, etc.) epitomizes a sensory memory intervention.

Short-term Memory (or Working Memory). The short-term memory component of human memory has a limited capacity in storing approximately 7 (–/+2) items of information for a duration of 15–30 seconds. Items can be retained in short-term memory by rehearsing them verbally [8]. An alternative model of short-term memory was developed at a later stage by Baddeley and Hitch [1974] known as working memory, arguing that the description of short-term memory provided by the multi-store model was overly simplistic, as mentioned above [12].

Long-term Memory. Memory theorists claim that the long-term memory could potentially have unlimited capacity with the main challenge lying in *how* able one is to access a memory and not *if* the memory per se is registered/available (i.e., accessibility over availability). In fact, Bahrick et al. (1975) investigated the capacity and duration of long-term memory in what they called Very Long Term Memory (VLTM), where they tested 392 participants of age 17 to 74, for memory retention periods of 2 weeks up to 57 years [15]. Tests included a free recall test (i.e., recall without any memory support), where participants were asked to recall the names of their classmates in a graduate class, a picture recognition test were participants reviewed a total of 50 pictures (mixed random pictures with pictures of their school albums), and a name recognition test for school friends. Results showed that participants who had already graduated for 15 years achieved a 90 % score in accurately identifying names and faces. 48 years after graduation, participants free recall of their classmates' names declined by about 60 % of the initial level, and cued-recall performance (using the album pictures as visual cues) declined more than 70 %. However, the decline of performance in these tasks over many years involves only the retrieval of names. The ability to identify names, faces and their associations remained unimpaired, although both recall tasks showed significant declines [15].

From a Neuroscience perspective, (conscious) human memory has been largely attributed to a particular brain area known as the "hippocampus" [79]. As such, long-term memory has been subdivided into a hippocampus-dependent memory component, known as declarative or explicit memory, and a hippocampus-independent memory component, known as non-declarative or implicit memory [86], as hypothesized in the multi-store model (see Figure 2.2). One of the earliest and most influential contributions in the domain of long-term memory was introduced by Tulving (1972). Tulving proposed the further distinction of declarative long-term memory into *episodic* and *semantic*, and that of implicit long-term memory into *procedural* memory, essentially extending the multi-store model [231] (see Figure 2.2).

Procedural memory is responsible for hosting skills and generally knowledge on how to do things. Procedural memory is also closely associated with habits, it is automatic and inarticulate, featuring both cognitive and motor activities [43]. For example, knowing how to open a door by grabbing and pressing down the handle is ascribed to our procedural memory, where information about motor skills and affordances (i.e., perceived action possibilities) is stored. Procedural memory is not declarative, typically does not involve conscious thought, and thus it falls under the implicit part of long-term memory, according to the multi-store model (see Figure 2.2). Nevertheless, procedural memory is outside the scope of this work and hence we do not dive into further details.

2.1.2 Other Memory Types

Over the years, multiple alternative classifications have been proposed to the established multi-store model. One of the most prominent classifications is based on the time order memories are stored for distinguishing between two broad categories in the long-term memory: **retrospective** and **prospective** memory components [29, 68]. On one hand retrospective memory typically encompasses the declarative branch of the long-term memory in the multi-store model (see Figure 2.2), including both semantic and episodic memory, though it can be explicit or implicit. On the other hand, prospective memory is where memories about the future are thought to be stored — described as "*remembering to remember*" type of memories. Prospective memories can be event-based (e.g., remember to turn off the lights when exiting the office) or time-based (e.g., remember to take the bus at 6pm), and can be triggered by a memory cue (e.g., see lights on, or see time almost 6pm). Nevertheless, the two memory components are not entirely independent, since an interplay between retrospective and prospective memory has been found [33].

From an alternative perspective, the ability to *learn* and *remember* a relationship between seemingly unrelated items or concepts (e.g., a song to a place, or the name of someone to the odour of a specific perfume) has been attributed to our **associative memory** [219], and it is known as "associative learning". A typical associative memory task involves testing participants' ability to recall pairs of unrelated items (e.g., face-name pairs) [156]. Perhaps the most eminent experimental demonstration of associative learning in mammals is the "Pavlov's dog"[2] experiment, where a seemingly irrelevant and neutral stimulus, such as a bell-ring, was associated with a biological potent stimulus, such as food, a phenomenon known in Psychology as "classical conditioning" [177]. In fact, evidence of associative memory intervention has been found in *episodic memory*, enabling the association among different aspects of spatio–temporal context (i.e., people, activity, location, and time) adhering to an episode [123]. Next, we describe in more detail how episodic memory works.

[2] `https://www.simplypsychology.org/pavlov.html`

2.1.3 Episodic Memory

The **episodic memory** component holds contextual information regarding *who*, *what*, *where*, and *when* of past experiences [226] — a summary of records of our life [48]. Once an episodic memory is formed, its information is said to be "*differentially accessible*", i.e., different sensory stimuli have different potential in triggering one recalling a particular personal past event. This is called "*episodic activation*" and describes the idea that in any episodic memory there is a pattern of activation that determines how its details are accessed [186]. The activation pattern for each episodic memory is determined by a range of factors, such as the goal of a past experience (e.g., trying to make it to the meeting on time), one's attention levels, the particular activity in which one was engaged (e.g., commuting), and one's affective state (i.e., summary of exhibited emotions over a period) during a past experience (e.g., feeling relieved arriving on time) [47]. Tulving describes this phenomenon as "encoding specificity", and posits that for a stimulus to be effective in triggering the recall of a past experience (i.e., **memory cue**), it should overlap with the memory trace of the memory to be retrieved [230]. Thus, memory cues that are as similar as possible to the initially encoded memory are particularly effective in triggering episodic memory recall (e.g., a picture featuring a beach one visited last summer helping one recall last summer vacations). The nature of episodic memory is dominated by visual imagery since images contain "*configural*" information [48]. Configural information originates from entities or objects represented in an image in relation to each other, revealing for example one's activity type or even intentions. This allows images to maximize the amount of information they contain. Furthermore, episodic memories have a "*field*" or an "*observer*" perspective [48]. Episodic memories with a field perspective maintain one's original point of view and thus, first-person view pictures captured with lifelogging wearable cameras have been found particularly effective in assisting one's recollections [111]. Instead, in an episodic memory with an observer perspective, one sees oneself in the memory when remembering [48] hence, imagery captured for example from infrastructure cameras may be more effective in assisting recall in this case. Notably, every time an episodic memory is recalled, it enters a malleable state, during which changes may occur to the episodic memory per se [231]. Finally, as the notion of temporal order is highly prevalent in episodic memories [48], one can typically remember more when past experiences are recalled in the order they took place [124]. A prevalent ethnographic method that leverages on this characteristic of the episodic memory is the Day Reconstruction Method (DRM), a retrospective self-report protocol that aims at increasing users' accuracy in reconstructing their experiences at the end of a day [124]. It does so by imposing a chronological order in reconstruction, thus providing a temporal context for the recall of each experience.

2.1.4 Semantic Memory

Similarly to episodic memory, **semantic memory** falls under the long-term memory category but holds information regarding events, facts and concepts, thus it is also known as *factual memory*. Particularly, it stores commonly shared general knowledge that is independent of personal experience or the context where it was acquired (e.g., water freezing point is 0° Celsius). Semantic memory is considered to be a highly structured network of concepts, words and images, capable for performing inferences, and inherently responsible for the use of language [231]. In contrast to episodic memory retrieval, when a semantic memory is retrieved its contents remain unchanged. Thus, semantic memory is considered as less prone to information loss when compared to episodic memory [231]. Interestingly though, semantic memory is connected with episodic memory in that we learn facts or concepts through experiencing, and in that respect episodic memory supports semantic memory. In fact, episodic and semantic memory components often operate in tandem, so that when either an episodic or semantic memory is recalled, both actions are registered as a new episodic memory [227]. However, over time, a transition from episodic to semantic memory occurs, during which an episodic memory gradually transforms into a semantic memory, reducing its association and sensitivity to particular events. As such, semantic memory has been associated with learning and performance, though it is believed that it is not directly involved during information acquisition (episodic is more prevalent), but rather in what can be retrieved after long time periods have elapsed [185]. From a Neuroscience perspective, semantic memory mainly activates the frontal and temporal cortices brain areas, whereas episodic memory activity is concentrated in the area of the hippocampus [155].

2.1.5 Memory Processes

Researchers have tried to improve our memory capacities with various techniques. These typically target one's short-term memory, trying to expand it with the help of constant training of memorizing numbers and/or card decks, or simply with verbal rehearsal (i.e., repeat information aloud to oneself) [118, 135]. However, when it comes to enhancing episodic and semantic memory recall, the challenge lies in improving the recall of specific episodic and semantic memories,

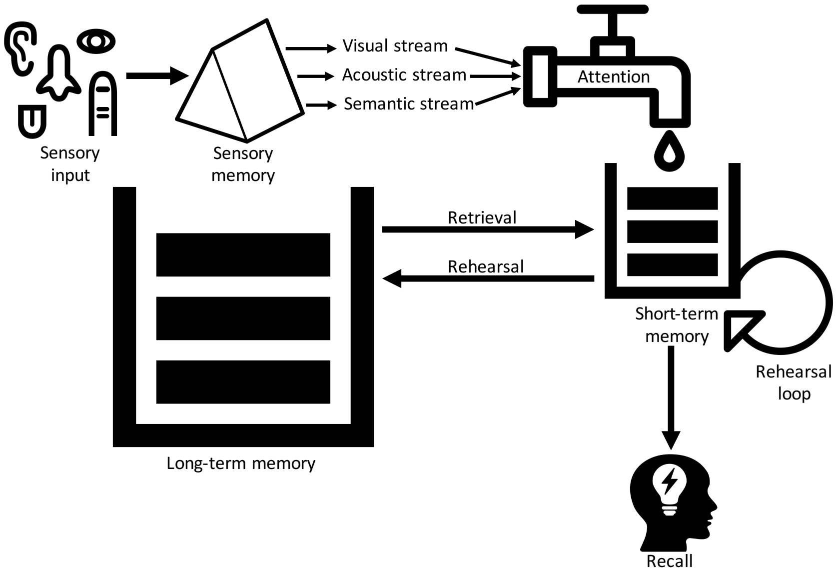


Figure 2.3. Encoding, storing and retrieving memories based on the multi-store model. Sensory input arrives one's sensory memory in the form of stimuli. Sensory memory encodes these stimuli in memory compatible forms such as visual, acoustic, semantic or tactile memories. Notice the role of attention reserves in regulating which memories reach our short-term memory. Then, subsequent rehearsals of those memories renders them as "permanently" stored in one's long-term memory.

not overall memory capacity, and thus a different approach should be considered. First though, we need to understand how memories are being stored and later recalled. Human memory encompasses five main processes: *memory encoding*, *memory consolidation*, *memory storage*, *memory recall* and eventually *forgetting*.

Encoding. Information arrives to our sensory memory via our five senses (i.e., our perception), but before it can move on to our short-memory system and further, it first needs to be encoded in a form that our memory can handle (see Figure 2.3). There are three primary memory forms[3] in which information can be encoded: *visual*, *acoustic* and *semantic*. Long-term memory, where episodic and semantic memory components reside, handles memory information of all three types. After perceived information is thus encoded into visual, acoustic and/or semantic in our sensory memory, it can be stored in our short-term memory (see Figure 2.3). Memories are assumed to be encoded and subsequently stored by a sparse population of neurons in the form of "*engrams*" or "*memory traces*" [122,

[3] A 4th memory form that of "*tactile memory*" has also been found [98].

148].

Consolidation. According to Atkinson and Schiffrin's multi-store model, our attention is critical for successfully registering incoming information to our short-term memory, and the more we rehearse this information (i.e., retrieve it from our short-term memory), the more likely it becomes "*permanently*" stored in our long-term memory [6], evoking a process known as consolidation. During consolidation, memories are stabilised by reducing any alterations in living neural tissue that corresponds to the engram of the memory under consolidation. [159] [92, 93, 99]. The consolidation process fires synchronously and repeatedly a group neurons (as a pattern of neural network activity, which can include the modification of synapses or creation of new ones among the neurons), making them permanently sensitized to each other, and thus more prone to activate together in the future (by a process called as long-term potentiation [26]). The process consists of synaptic consolidation (within the first few hours after encoding) and system consolidation (over a period of weeks to years). Memory consolidation processes are known to manifest during sleep and are particularly beneficial for learning [82].

Storage. After a memory has been successfully encoded and sufficiently consolidated, it is presumed to be stored to the long-term memory in the form of engrams [122, 148]. As information flows from sensory to short-term and finally to long-term memory, it gets filtered for avoiding information overload. The more a memory is recalled (e.g., through rehearsal), the more likely it is to be retained in the long-term memory because of triggering subsequent consolidation processes (see Figure 2.3). However, memories are innately dynamic, since the very activation of a pattern of neural activity due to recall may modify the engram itself, resulting in registering new information to the memory (modifying how the original life experience is remembered) [158].

Recall. For successfully recalling (or retrieving) a past experience or prior knowledge, our memory system utilizes engrams (i.e., memory traces) and encoded contextual information that arrives in our short-term memory as input from our sensory memory and via our perception (see Figure 2.3). The contextual information may arrive explicitly (intentionally – e.g., one reviewing one's family album for reminiscing) or implicitly (unintentionally – e.g., the smell of bee-wax candles reminding one of past Christmas celebrations). This contextual information is basis of what is known in Psychology as **memory cue**. Contextual information is what we can perceive with our five senses and can be anything from pictures, locations, sounds to smells, tastes, and haptic feedback. For example, in order to support patients suffering from amnesia, one can explicitly target the triggering of their episodic memory by providing such relevant contextual information in the form of browsing photo albums. This process of replaying previously recorded relevant memory cues for triggering one's episodic memory recall is called *cued recall* [35]. Cued recall can help elicit episodic and semantic memories for healthy individuals too. For example, one viewing a picture from one's childhood could trigger the episodic recall of a long-stored memory from that time (e.g., swimming at the sea with parents during summer vacations).

Forgetting. However, it is often the case that the recall of a memory may fail. Forgetting is considered as a temporary or permanent inability to recall a piece of information previously stored in the brain, due to a mismatch between memory cues and the encoding of the information (incorrectly or incompletely encoded memories, and/or problems with the recall process). However, the memory may still be stored but rendered inaccessible [225]. This may be due to a memory disorder (e.g., Dementia) or simply due to natural loss of information over time. Memory attenuation over time has been described by a plethora of so-called "forgetting functions" that attempt to best approximate the empirical Ebbinghaus' forgetting curve [242] (see Figure 2.1). In that respect, perhaps the most prominent forgetting function is the Wickelgren's power law [244]:

$$m = \lambda(1 + \beta t)^{-\psi}, \tag{2.1}$$

where m is memory strength, and t is time elapsed since encoding (i.e., the retention interval). The equation has three parameters: λ is the state of long-term memory at $t = 0$ (i.e., right after the encoding), ψ is the rate of forgetting, and β is a scaling parameter [248]. Forgetting is typically viewed as an unwelcome side-effect of memory rather than a memory process. Nevertheless, forgetting may be critical for our memory in filtering information overload, recovering from negative experiences [201], forming new memories and acquiring new knowledge [55]. Schacter has rigorously described forgetting in "The Seven Sins of Memory", namely *transience*, *absent-mindedness*, *blocking*, *misattribution*, *suggestibility*, *bias*, and *persistence* [202]. We briefly describe them next:

1. **Transience** stresses the fact that memory attenuation can vary from naturally gradual (i.e., long-term) to quite rapid (i.e., short-term). Typically, memories that are not retrieved or rehearsed may slowly dissipate over time.
2. **Absent-Mindedness** describes the lack of attention during encoding (i.e., "shallow encoding") or retrieval, occurring when information is superficially processed. Absent-mindedness during retrieval is expressed as forgetting to carry out a particular task or function, and thus are typically referred as failures of the prospective memory.

3. **Blocking** refers to the phenomenon during which a deeply encoded memory is temporarily inaccessible (i.e., retrieval block), and is known to manifest both in episodic and semantic memory. Tip-of-the-tongue (TOT) state is the most prevalent example of blocking and is believed to occur due to the provision of memory cues that are related to a previously recalled memory (i.e., "part-set cuing" effect). Blocking is known for exacerbating with ageing.
4. **Misattribution** highlights the situations when a memory is accessible but is erroneously ascribed to an incorrect context (i.e., place, time, person, and activity). An explanation may be that one solely relies on the general semantic features of the incorrectly recalled memory.
5. **Suggestibility** refers to the tendency to incorporate information by others during recall (i.e., misleading questions), and is closely related to misattribution. In fact, it is possible to implant false episodic memories by utilizing diverse suggestive procedures (e.g., hypnosis).
6. **Bias** postulates that memory encoding and retrieval are highly susceptible to pre-existing knowledge and beliefs. This happens due to the natural tendency for showcasing a consistency between past and present attitudes, beliefs, and feelings (i.e., consistency/retrospective bias — see also Festinger's theory of "Cognitive Dissonance" [76]).
7. **Persistence** involves the recall of an episode that one would prefer to forget (e.g., a traumatic experience). In fact, traumatic memories can often be more disrupting than forgetting per se, due to their emotional bearing. Emotion has been associated with vivid recollections due to the intervention of the "amygdala", a brain area responsible for regulating emotion and emotional memories.

In this thesis, we utilize the memory triggering potential of contextual information (e.g., pictures, videos, text, etc.) for the selection and generation of memory cues. Memory cues are then replayed at random or predefined moments on one's personal devices (e.g., smartphone) for augmenting one's memory about a past experience or prior knowledge. In particular, our work draws on intentionally and repeatedly prompting memory recall processes about a targeted memory (i.e., memory subjected to augmentation), for implicitly evoking memory consolidation processes corresponding to the targeted memory. As we saw above, repeated instances of memory recall-consolidation processes lead to permanently storing a memory. This phenomenon comprises the essence of *cued recall*, the theoretical underpinning of this work. Due to the potential of (mobile) cue-based memory augmentation in forming strong memories, we believe that the memory augmentation effect will persist through time, even without further technological support.

2.2 Lifelogging

Lifelogging is the practice of continuous capture of (multi–modal) data streams that characterize a life experience. The data streams are typically stored in a system that people can use for reviewing past experiences and episodes of their everyday lives, for reminiscing, self-reflection, or planning future actions and alter/sustain behaviours and habits [93, 102]. Lifelogging as an idea was first introduced by the American engineer Vannevar Busch in 1945, when he proposed the "*Memex*" [84] concept as a device that one would store individually one's books, records, and communications, with the purpose of consulting it later for enlarging and supplementing one's memory. With the advent of technology, especially the miniaturization of capture hardware such as cameras and the dramatic capacity increase of cheap storage, lifelogging became increasingly popular. However, there is no universal consensus on what exactly lifelogging is, as it turns out to be a combination of quantified-self[4] analytics, life-blogs (i.e., documenting life experiences), personal digital memories (i.e., personal multimedia) and others. Nevertheless, a more formal definition of lifelogging is the following:

> "*Lifelogging is a form of pervasive computing, consisting of a unified digital record of the totality of individual's experiences, captured multimodally through digital sensors and stored permanently as a personal multimedia archive*" [65].

Lifelogging is intended to be ubiquitous, recording continuously and passively a wide range of aspects about one's daily life through some sort of lifelogging device. Thus, a **lifelogging device** can be wearable cameras (e.g., the Narrative Clip), physiological monitoring devices (e.g., the Empatica E4 wristband), fitness trackers (e.g., Fitbit), a modern mobile device, infrastructure cameras and other sorts of devices that capture aspects and activities of our daily lives. Captured data from different sources are then typically processed and synced to form personal digital multimedia records, known as **lifelogs**, and are permanently stored. Besides capture and storage, a lifelogging system encompasses all the necessary processes for processing and retrieving a lifelog. The volume and variety of data

[4]A movement for utilizing technology for acquiring data that describes aspects of one's daily life in terms of inputs (food consumed, air pollution, etc.), states (mood, heart rate, etc.) and mental or physical performance (chess matches won, distance run, etc.).

that comprise a lifelog can be enormous, and thus the development of lifelogging systems is particularly challenging [83, 93, 255]. For example, often the manual retrieval of the desired information may be impossible due to the volume of data stored. Hence, **lifelogging user interfaces (LUIs)** are employed for facilitating the interaction between lifelogging systems and users. LUIs are responsible for supporting the synergy between human memory and a lifelog by providing the necessary searching and presentation functionalities [93].

Predominantly, the motivation for lifelogging concerns the benefits for one's memory, and has been prominently summarized by Sellen and Whittaker [2010] in the "**Five Rs**" [210]. We briefly present them next:

1. **Recollecting** describes the potential of lifelogging in assisting one to "re-live" a past experience for utilitarian reasons (e.g., remembering what was discussed in a meeting).
2. **Reminiscing** describes the notion of recalling a past experience for re-exhibiting past emotions (e.g., recalling a trip to Hawaii).
3. **Retrieving** underscores the potential of lifelogging to locate a specific piece of information without necessarily recalling the whole past experience (e.g., retrieving an old e-mail).
4. **Reflecting** highlights the use of lifelogging for introspection through recalling a past experience and analysing past behaviours (e.g., recall oneself before quitting smoking).
5. **Remembering intentions** describes the value of lifelogging in assisting our prospective memory (memory about the future) and goal-tracking (e.g., remember to do the groceries).

In fact, lifelogging can benefit a large spectrum of different application scenarios, other than supporting human memory, such as healthcare, learning, behaviour change, productivity, leisure and others [93], but in the scope of this work, we elaborate on the potential of lifelogging to augment human memory recall [102].

Justifiably, one would expect that lifelogging is the best answer to improving recall of past experiences. In fact, given today's abundance of cheap storage and the ubiquity of capture devices, we could literally keep capturing everything forever. However, especially when it comes to pictures (and videos), lifelogging produces a sheer volume of data that cannot simply be browsed in the way traditional photo albums could. Not surprisingly, a considerable portion of lifelogging data ends up never being reviewed [179, 210]. Also, despite the use of filtering algorithms and summarization techniques (e.g., clustering pictures by spatial and/or temporal affinity), often captured imagery does not vary significantly, particularly for daily routines (e.g., when commuting or at the office), and hence rendering the review process a rather monotonous task. In this work, we explore ways in which contextual information as captured via additional wearable devices (e.g., physiological measuring wristbands) can help identify "*moments of significance*" (e.g., high arousal), and thus result in the selection of more effective memory cues. For example, physiological responses could indicate pictures of increased interest, and hence memory value, as we will see in Chapter 8.

2.2.1 Lifelogs as Digital Mementos

In the struggle for protecting memories from fading away, people have for long collected items that represent past memorable experiences. Nowadays, a digital picture, or any other memory-supportive contextual information (e.g., video, sound, location, etc.), serves as a "*digital memento*" (i.e., digital souvenir) of a past experience — an object given, collected or kept as a reminder of a person, place or event [180]. Digital mementos are usually captured via technology (e.g., taking a picture with a smartphone), as opposed to physical mementos, which are collected (e.g., buying a postcard). In both cases, the purpose of mementos is to assist remembering at a later stage. Besides, the act of collecting mementos is what distinguishes an event one wishes to remember, from a mundane or indifferent event [243]. However, digital mementos bear some significant disadvantages when contrasted to physical ones. In general, digital mementos are considered as less salient when compared to physical ones, since they are innately intangible, thus still less integrated in everyday life, and rather "imprisoned" in digital storage [179]. The ease at which one manipulates digital mementos (e.g., copy, delete and modify) renders them ephemeral, often impersonal to convey the richness of a past experience (especially when captured automatically), and tedious to access, as opposed to their physical counterparts. For example, reminiscing one's school years when holding and flipping through an old and dusty photo album feels more natural and intimate than browsing through digital images on a computer screen. Nevertheless, digital mementos can easily be collected at large volumes and with minimum cost, they can easily be shared, edited and dynamically displayed, and they can be stored virtually forever [233]. Consequently, lifelogging has largely automated the process of collecting digital mementos in the form of *lifelogs*.

2.2.2 Creating and Using a Lifelog

There is no universal consensus over the processes that pertain to lifelogging, since different lifelogging systems appear to implement different processes or stages for producing and using a lifelog. Nevertheless, the following processes have been identified in literature and are presented here as the most frequently reported:

Capture. The first step is the collection of raw data (e.g., pictures, videos, locations, etc.) that correspond to a life experience or event, commonly with the use of a lifelogging device [93, 184, 209]. Data capture typically occurs automatically, continuously, and transparently, without the user's intervention. Often, captured data has to be readily available, and hence the employed lifelogging devices are continuously streaming captured data to a central storage for real-time analysis. Application scenarios based on readily available data include supporting one's prospective memory (i.e., memory about the future), or simply logging for security reasons (e.g., an car dash cam). Alternatively, captured data may be opportunistically offloaded to the central storage of a lifelogging system (e.g., when Wi-Fi connection is available). Evidently, the captured data can be immensely heterogeneous (e.g., visual, audible, spatial, physiological, activity type, etc.), since it originates from the collection of multi-modal data streams from a plethora of sources, such as sensors (e.g., portable or fixed cameras, microphones, wearable and environmental sensors, etc.), services (e.g., stored e-mails on a mail server), (mobile) applications (e.g., chat transcripts), and networks (e.g., social media posts and activity), typically by accessing an Application Programming Interface (API) [93].

Synchronization. The great heterogeneity that characterizes captured data renders the synchronization process as the necessary next step. At this stage, captured data may be re-formatted to match formats employed by the lifelogging system, and are temporally and spatially synchronized, particularly when it comes to handling multi-modal data streams [93]. Then, data is filtered and clustered. Ideally, data synchronization is immediate and automatic, but often a delay is entailed due to data offloading employed by lifelogging devices. Synchronization can be instantaneous and automatic when the lifelogging device is in constant communication with the central storage.

Segmentation. Once data has been synchronized, (temporally, spatially, etc.), related data chunks are combined and restructured for forming meaningful data segments, usually in the form of events or episodes, which can be utilized as discrete "*units of retrieval*" [93, 184]. Retrieval units, in the form of episodes, should be viewed as single "*semantic units*" that can be ordered based on temporal or spatial criteria [102]. These episodes represent the actual life experience of the user, and are linked to each other in a fashion that facilitates the reconstruction of the actual experience, in a similar way as human memory does (e.g., in temporal order) [231]. These episodes can be utilized in subsequent analyses for determining their uniqueness and regularity, thus extracting behaviour patterns for supporting additional purposes (e.g., for goal-tracking and self-reflection [210]).

Enrichment. At this stage, episodes are semantically and emotionally enriched for creating a sort of narrative or story. The enrichment takes place by labelling the episodes with meaningful meta-data, such as occasion, place name, people, activity, feelings, and others, as it is required for enhancing the usefulness of the episodes that constitute a lifelog [93, 94, 150]. The information for the enrichment is derived from additional sources, such as data from diverse sensors, information from temporally adjacent episodes, APIs, semantic analysis, face recognition, and others, so that new meaning can be attributed to an episode. This process can be automatic or manual, with users reviewing and annotating their data, though only a few ever retrospectively edit, manage, or curate their data (e.g., deleting duplicates) [70]. Depending on the application of lifelogging per se, a rich narrative summary could be constructed for supporting keyword text search. The enrichment process is thus critical from an indexing perspective, since the elicited labels (i.e., meta-data annotations) can be used for retrieval in the next stage.

Retrieval. At this stage, the lifelog related to a life experience has been created and is ready for retrieval, typically via a search engine [93, 94, 150, 184]. We need to stress that the successful retrieval of a lifelog is heavily depending on the nature of the circumstances under which a lifelog is about to be used. This implies that understanding user's intentions in how lifelogs are accessed and used, in different (lifelogging) application scenarios, is of utmost importance. Based on the lifelogging application scenario per se, a Lifelogging User Interface (LUI) is responsible for providing the appropriate retrieval functionalities and controls.

Interaction. A Lifelogging User Interface (LUI) enables the communication between the user and the lifelogs [94, 184]. For doing so, a LUI encompasses the necessary graphical elements for furnishing input and output control, and thus facilitating overall interaction. Typically, a LUI hosts search and retrieval options and functionalities for actuating the lifelog retrieval process. In particular, lifelogs can be retrieved, presented, and visualized, using appropriate modes of presentation, visualization, and information prioritization. For example, when one interacts with a LUI for supporting one's memory, the following user interaction pattern is followed: (a) *overview*, (b) *zoom and filter*, (c) *details* *on demand*. Depending on the application of lifelogging per se, the input controls of a LUIs usually support assisted query formulation, engaging storytelling, summarization, visualization, and the delivery of information that calls for action (e.g. recommendations or reminders). The interaction with a LUI can be multi-modal (e.g., on personal computers, tablets, smartphones, smartwatches, ambient displays, etc.) for rendering a lifelog more accessible and pervasive. Ideally, the interaction with a LUI should be the only "*touchpoint*" that requires the active participation of the user for retrieving a lifelog. However, not all lifelogging systems implement the aforementioned lifelogging processes, or even if they do, they may require user intervention. When a lifelogging system supports the prospective memory (i.e., remind one to drink more water), the interaction is usually initiated by the system via a LUI (e.g., display reminder on one's smartwatch). As such, the lifelog is retrieved and presented seemingly unexpectedly for the user, based on some contextual criteria (e.g., time, location, etc.).

2.2.3 SenseCam: Lifelogging in the Service of Human Memory

As we have seen in Section 2.1.3, episodic memory is innately visual, and thus visual information in the form of **memory cues** (i.e., videos/pictures) can be particularly effective in triggering episodic recollections. Perhaps the most prominent lifelogging application for supporting memory recall is SenseCam (see Figure 2.4), a small wearable camera that can be worn on a cord around one's neck, capable of automatically capturing an image every 30 seconds (i.e., time-driven capture), or every time it detects ambient light changes, temperature changes, or movement (i.e., event-driven capture) [111]. Developed by Microsoft and first introduced in 2004, SenseCam[5] is equipped with wide-angle (fish-eye) lens capturing most of what the wearer sees, including social interactions, while alleviating one from the burden of manually taking a picture, but without providing a preview, unlike a typical digital camera or smartphone. Other features include the explicit capture, triggered manually by the user, same as with a typical camera, and a privacy button that suspends image capture for four minutes. The latest SenseCam version (v2.3) captures images at a VGA (Video Graphics Array) resolution of (640x480 pixels) and stores them in ".jpg" format, on an internal flash memory of 1 GB, which typically can store over 30,000 images. Along with the images, SenseCam also records meta-data such as image timestamps, sensor measurements (i.e., light intensity, temperature, and 3-axis acceleration), and GPS traces. The images recorded by SenseCam can be download onto a com-

[5] `https://www.microsoft.com/en-us/research/project/sensecam/`

puter and can be browsed using the SenseCam Image Viewer[6], which displays them in a sequence, at a slow (2 images/second) or fast (10 images/second) pace.

Figure 2.4. Microsoft's SenseCam.

The presentation of images individually and in temporal order leverages on the "personal" movie paradigm, greatly supporting the recall of autobiographical memories, since the notion of time is central for episodic memory [48]. Therefore, the SenseCam was initially utilized as a retrospective memory aid for patients with impaired memory, such as those suffering from Alzheimer, Dementia, and Amnesia [111]. Interestingly, it was found that reviewing the images taken can significantly improve subsequent recall performance not only in memory-impaired users [23, 144, 250], but healthy users too [209]. The continuous collection of one's daily encounters renders highly likely the possibility that a memory cue, triggering a memory one struggles to recall, is in fact captured within the temporally-ordered image sequence of that day. Typically, SenseCam's effectiveness in supporting memory recall has been ascribed to the end-of-day review of images corresponding to salient episodes of that day [102]. Hodges et al., believe that the SenseCam images provide effective **memory cues** to past personal experiences, eliciting thoughts, feelings and emotions, clinging to the

[6]`https://www.microsoft.com/en-us/download/details.aspx?id=52298`

original experience, and thus evoking richer recollections [111]. However, most of the studies using SenseCam, or SenseCam-like approaches, focus entirely on augmenting the recall of experiences per se, while ignoring *how* and *if* emotions or feelings are recalled and re-experienced. In Chapter 4, we present a study that focused particularly on how mobile users recall or recognize emotion from their facial expressions and their surroundings, as depicted in automatically captured "selfies" throughout the day [165].

2.2.4 Design Principles for Lifelogging User Interfaces

A Lifelogging User Interface (LUI) serves as the mediator for accessing and managing lifelogs. Hence, since the interaction between the lifelogger and the lifelog is solely supported by the LUI, the LUI's design is essential. The following design principles for LUIs have been elicited from literature, and are presented in this section. The intention here is to provide insight and assistance for the design of LUIs that best support human memory through lifelogging [102]. Ultimately, the elicited design principles also contribute to the design and development of future pervasive memory augmentation systems, as shown in Chapter 9.

Expressiveness and Meaningfulness of Lifelogs

The intangible nature of lifelogs, as digital mementos, prevents them from being directly situated in the physical space — such as physical mementos are — and hence rely on technological interventions for being accessed (e.g., a video from last summer vacations playing on a domestic display) [179]. Hence, LUIs should be designed in a way that maximizes the expressiveness of lifelog content. This can be achieved by underscoring the highly experiential and affective aspects of lifelogging for visualizing a lifelog in a way that appears more personal and intimate to the user [36]. The thoughtful use of color, transitions and animations establishes an engaging and enjoyable user experience, and in turn, may increase the perceived meaningfulness of a lifelog [179, 252].

> **Principle 1:** Utilize the experiential and affective characteristics of lifelogging for increasing the perceived expressiveness and meaningfulness of lifelogs.

Quick Overview and Prioritization

During lifelogging, the automatic collection of immense volumes of data entails the frequent production of lifelogs that may hold little if any memory value [210].

This can be attributed to the always-on nature of lifelogging, sporadically capturing memorable experiences, but rather often mundane activities too [199]. In fact, most of the captured lifelogging content is apropos to irrelevant, repetitive and monotonous activities (e.g., commuting) [36]. However, when users interact with LUIs, they are particularly interested in obtaining an overview of the retrieved lifelogs, and then quickly fine-tune their scope to relevant episodes [114]. Thus, LUIs should first group episodes by affinity, and subsequently rank them by utilizing custom ranking criteria that prioritize more relevant and important episodes (e.g., retrieve five most intense running experiences) [114, 233].

Principle 2: Provide a quick overview of relevant episodes, rank them by prioritizing the most relevant, and apply prominent visualization techniques for ensuring they stand out.

Flexible Navigation and Exploration

The ever-growing lifelog repositories pose an unparalleled burden in browsing through or searching for lifelogs, often entailing significant cognitive overhead for the user [36, 114]. A good design practice for a LUI is the presentation of lifelogs in episodes, while visualizing any relations with each other [36, 94, 233]. The episode presentation analogy enables browsing and search by utilizing useful retrieval units, in a similar way as human memory functions [231]. However, there is no uniform template for "calibrating" the scope of an episode (i.e., retrieval units), as it seems to depend on the nature of the lifelogging application per se [93]. For assisting memory, it is often the case that only very specific information may be required (i.e., sub-episode), or instead, a collection of lifelogs (i.e., supra-episode) [94, 150, 210, 243]. Thus, an episode may not always employ the necessary level of abstraction for presenting a lifelog that supports memory.

Principle 3: Use entire collections of episodes as retrieval units, while supporting flexible navigation and exploration of the corresponding lifelogs.

Adaptation of Lifelog Content

The role of a LUI is to facilitate the synergy between human memory and a lifelog for assisting memory recall [127, 150, 243]. The interaction with lifelogs through a LUI can help one track and retrieve missing specks of past life experiences (i.e., **memory cues**), and thus effectively remove bias by "repairing" distorted or malformed memories [114]. For effectively supporting memory recall, LUIs should present and visualize effective memory cues apropos to user's intentions [93]. The previously mentioned "Five Rs" provide a good roadmap to approximating user's intentions when utilizing lifelogging for supporting memory recall [93, 210]:

1. For *recollecting*, LUIs should present lifelogs from both 1st and 3rd person perspectives (if available), in assisting one "re-living" a past experience, by leveraging on the visual nature of episodic memory [231]. A LUI should incorporate interaction techniques that enable accurate assisted query formulation, even when the query is incomplete [66, 93].
2. For *reminiscing*, often groups (i.e., family, friends, etc.) engage in a social activity during which storytelling or sharing of life experiences take place [147]. A LUI should thus support lifelog sharing with others and collaborative search through visual modalities.
3. For *retrieving* information, LUIs should support the user in the query generation process, which depends on user's ability to recall the initial query keywords for generating an effective query [93]. In retrieving, a user seeks to recover a particular piece of (atomic) information (e.g., a name, an address, etc.), and hence the formulated query will define the type of knowledge to be retrieved along with the employed visualization (e.g., a map for visualizing an address).
4. For *reflection*, LUIs should support lifelog summarization, knowledge inference, and the presentation of meta-data (through pattern analysis and discovery), in ways that might surprise, provoke and educate based on knowledge that may not be directly evident [117]. Thus, lifelogs should be visualized by utilizing pertinent interface metaphors (e.g., timelines, maps, graphs, etc.), while supporting click-through analysis of the lifelog's underlying data, for drilled-down reflection [93].
5. For *remembering intentions*, LUIs should focus on delivering lifelogs as memory cues or reminders for supporting the prospective memory (i.e., remembering future plans) based on time or other contextual criteria (e.g., location, activity, co-presence, etc.) [94]. As such, the in-time delivery of a memory cue may be critical for successfully assisting prospective memory. Consequently, a LUI should support multi-modal interaction (e.g., personal computer, smartphone, ambient display, smartwatch, etc.), utilize the adequate modality based on user's context and preferences (i.e., visual, audible, tactile or even olfactory [63]), and based on these, select the appropriate lifelog as a memory cue for user's prospective memory.

Principle 4: Adapt the presentation and visualization of lifelogs to the purpose of recalling while supporting flexible query generation.

Contextualization of Interaction and Visualization

Most of the approaches for the visualization of lifelogs assume a generic LUI for every type of lifelog content [114]. However, the range of potential users of lifelog content is diverse in terms of technological shrewdness, skills, and experience [36]. For example, teenagers are usually more technologically adept than elders are. Furthermore, lifelogging can be applied to a large set of application scenarios, with each one requiring specific lifelog visualization and interaction techniques. For instance a LUI should employ different visualization when presenting lifelogs from a last jogging session for encouraging one to exercise more (i.e., for reflecting), as opposed to when presenting lifelogs from last year's summer vacations (i.e., for reminiscing). In practice, it is particularly difficult for a LUI to satisfy all these requirements, however precious insights can be gained by the thoughtful elicitation of application scenarios and use cases during the design of a LUI [36, 93].

Principle 5: Consider user characteristics and use case application scenarios for contextualizing lifelog visualization and interaction.

Speed and Accuracy Throttling

Often, important lifelog meta-data, in the form of labels or annotations (e.g., time, place, person, activity, etc.), may not be available, and hence inhibiting a quick and successful retrieval of lifelogs [210]. Moreover, users seek to recall more than just facts, and tend to prefer speed of retrieval over accuracy of the retrieved lifelogs, depending on how urgent is the demand for recall [127]. However, a lifelog repository can potentially host information that spans over one's lifetime [36], rendering fast access, search and retrieval, a real conundrum. Thus, a LUI should feature efficient navigation controls for quickly exploring retrieved lifelogs by providing an overview first, zoom and details on demand. LUIs should display relationships between related lifelogs and provide comparisons and summaries based on time periods and semantic concepts, so that users can interpret episodes over time, even when lacking meta-data [114, 243].

Principle 6: Provide efficient navigation and exploration controls for satisfying speed and accuracy of retrieval on demand.

Multi-Modal Interaction

The ubiquity of personal devices in various screen sizes and capabilities, has greatly influenced the design of modern graphical user interfaces (GUIs). Consequently, LUIs should support seamless multi-modal interaction with lifelogs on the tiny screen of a smartwatch, to a huge domestic display (e.g., a smart TV) for essentially supporting memory recall [56]. Thus, the range of potential application scenarios for ligelogging has greatly increased, encompassing diverse contexts and numerous use cases. For example, one can quickly glimpse on a lifelog appearing on one's smartwatch when at the bus stop, or review last skiing video recorded on one's smartphone when on a ski lift [75]. Evidently, not all lifelogs are retrieved and reviewed in the same context and in the same frequency [70]. Thus, LUIs should be designed with multi-modality in mind for maximizing lifelog availability and presentation based on context of use [252].

Principle 7: Consider the features of the employed interaction modality for maximizing the capability of LUI in presenting and visualizing lifelogs.

Reinforcement of Synergy

Information related to a particular episode is encoded and stored in human memory as a network of interrelated facets for every past experience (e.g., emotions, objects, people, events, places, etc.) [146]. During the recall process these facets serve as index entries for a given episode. Prompting a single facet may be enough for triggering the recall of an entire past experience. During the retrieval process in lifelogging, these facets are prompted by **memory cues** presented and visualized as part of a lifelog [36, 40]. Often, a LUI may not be able to prompt the entirety of facets describing a past experience, simply because the corresponding memory cues have not been holistically captured [36]. For example, a selfie as a memory cue, may contain one's facial expressions and thus exhibited emotions (i.e., emotional facet) but not other people (i.e., co-presence). There is no universal consensus on which type of memory cues (corresponding to facets) should be visualized in LUIs, however four types are more prevalent: i) *activity*, *event* or *action* (what one was doing), ii) *location* (where one was), iii) *time* (when it happened), and iv) *co-presence* (who one was with) [39, 85, 127, 184, 233]. Other types of memory cues are less frequently reported but appear to be relevant: feelings or emotions (how one felt) [144], thoughts (what one thought) [144], objects (digital and physical) [126, 180], physiological responses (e.g., heart rate, galvanic skin response, etc.) [150, 168, 200], and sound (e.g., ambient sound and conversation recordings) [109, 178]. Not all facets are equally capable in triggering memories through the presentation of the corresponding memory cues [127], but *activity*, *place*, and *people* are considered the most effective for synergistically supporting the recall process via a LUI [210].

Principle 8: Visualize memory cues for prompting activity, location, and co-presence memory facets that will synergistically support memory recall.

2.2.5 Lifelogging as Human Augmentation

Already by 1965 (the year of the first space-walk), people envisioned a future where humans are linked to computers in a symbiotic manner for enhanced memory and cognition. The 1965 Sunday comic strip "*Our New Age*" stated: "*By 2016, man's intelligence and intellect will be able to be increased by drugs and by linking human brains directly to computers!*"[7]. Under-explored, equally like outer-space, the study of human brain gave slowly birth to Neurotechnology — in the early 1970s — proposing the use of electroencephalography (EEG) as a new method for directly linking brain activity with computers [235]. Interestingly, to date, computer systems and the human brain have already formed a basic — unidirectional — communication channel through Brain-Computer Interfaces (BCIs). Disabled people can now learn to control robotic limbs by the sheer power of their mind [249], stroke survivors can manipulate virtual limbs in virtual reality environments [236], up to brain-controlled computer games designed for entertainment [204]. Undoubtedly, BCIs have driven a revolution in the areas of assistive and rehabilitative technologies, increasing quality of life for people with kinetic or other impairments. In fact, brain monitoring and sensing technologies, such as BCIs, may have the potential to revolutionize the entire spectrum of our cognitive processes, and particularly our memory, when incorporated with context-capturing and technology-driven practices, such as lifelogging and the quantified-self movement.

For Neuroscience and Psychology, enhancing human memory has been the epitome for a multitude of renown studies, while providing ample ground for heated discourse. The focal point of the debate is primarily the level of invasiveness of the employed cognitive intervention, with some arguing for brain implants (typically neuroscientists), while others prefer non-invasive methods such as memory training (typically psychologists), while electric brain stimulation emerges as a middle-ground [74]. In fact, a very recent breakthrough by Hirschberg et al., showcased a neural probe (i.e., implant) prototype [110] ca-

[7] `https://www.smithsonianmag.com/history/sunday-funnies-blast-off-into-the-space-age-81559551/`

pable of increasing performance in memory tests with human subjects by up to 30 %[8]. So far, similar invasive approaches for improving memory or repairing fallible memory, were only tested in non-human subjects (e.g., Long-Evans rats) [21].

As we previously discussed, the always-on nature of lifelogging may provide a viable and non-invasive alternative in augmenting our memory, via the actuation of modern, cutting-edge technology. In fact, the increasing sensing capabilities of wearable technologies such as BCIs, "smart" wristbands that monitor physiological responses (e.g, heart rate, electro-dermal activity, etc.), the advent of Virtual Reality (VR) and Augmented Reality (AR) headsets (e.g., Microsoft HoloLens[9], the newly announced Magic Leap One[10], etc.), and the ubiquity of ambient displays, present an unparalleled opportunity to holistically and significantly improve the way we encode and retrieve memories. In particular, the continuous collection of memory-supportive contextual information (e.g., activity, location, co-presence, etc.), with the unobtrusive monitoring of our cognitive and physiological responses can facilitate the in-time (e.g., when one is bored [182]) and ambient presentation of effective **memory cues**, that greatly help us remember a past experience (episodic memory), prior knowledge (semantic memory), or an upcoming event (prospective memory).

In this work, we attempt to take the next step towards utilizing lifelogging as a non-invasive approach to human memory augmentation. We start with the automatic and unobtrusive capture of contextual information from diverse sources for creating effective memory cues that facilitate serendipitous memory recall. For example, in Chapter 4, we investigate how event-driven picture capture can provide a good source of memory cues for augmenting the recall of episodic memories for inferring emotion [165]. Next, we utilize the memory augmentation potential of our approach by trialling it in various application scenarios for measuring UX levels, or helping one recall a past meeting (Chapters 5 and 7, respectively). In this way, we showcase how memory augmentation through lifelogging can be particular useful for a wide range of domains (e.g., mobile application design, well-being, automotive, workplace, etc.), surpassing the narrow boundaries of the self [165, 166, 169]. Nevertheless, as we acquire the potential to gather increasingly large lifelogging datasets, managing them becomes a significant challenge. Luckily though, the proliferation of data that describe

[8]https://www.newscientist.com/article/2153034-brain-implant-boosts-human-memory-by-mimicking-how-we-learn/

[9]https://www.microsoft.com/en-us/hololens

[10]https://techcrunch.com/2017/12/20/magic-leap-shows-off-its-magic-leap-one-creators-edition-ar-headset-shipping-in-2018/

one's physiological responses and emotional states could perhaps indicate significant or worth-remembering moments, increasing perceived usefulness and pervasiveness of lifelogging. In Chapter 8, we highlight the potential of physiological responses in memory cue selection and delivery, for further augmenting memory recall with lifelogging [153, 168]. However, we believe that in order to realize the full potential of lifelogging in holistically augmenting human memory throughout daily life, it is necessary that the so-called "*cognitive gap*" between the human (i.e., user) and the machine (i.e., system, or technology in general) should close [170].

2.2.6 Beyond Lifelogging

Over the millennia, the human brain has evolved to excel in collecting (perception) and processing information (cognition) in an incredibly efficient manner. Essentially, the way our brain has evolved and has been structured is what makes us so different from all other animal species, many of which have significantly larger brains with even higher number of synapses [207]. For the human brain to continue evolving, certain organs may have to become bigger, including the human brain per se. For example, for processing more information, wider synapses are needed, resulting in a demand for greater in-brain blood flow and in turn a larger heart. However, even if the human brain and its supporting organs and networks will eventually grow in size, further evolution will certainly face limits imposed by laws of Physics, as well as diminishing efficiency after a certain threshold of brain size increase [112]. Nevertheless, these changes cannot occur naturally and in a timely manner for satisfying the frenetically increasing demands in information processing of the today's era. In fact, lifelogging, could be considered a plausible culprit for aggravating information overload in a future uptake, simply due to the utter volume of data it entails [64].

Evidently, the way people collect and process information has always been influenced by technology. For example, the introduction of stone-headed javelins during hunting would have greatly altered decision making and strategic thinking of prehistoric humans. Similarly, the introduction of smartphones has greatly influenced the way we seek and consume information, highly disrupting the entire spectrum of our cognitive processes (i.e., attention, memory, learning, decision making, problem solving, etc.). Albeit technology so far had a beneficial role on how we perceive and process the world around us, in recent years it has undertaken a rather disruptive and double-edged role. The era of ubiquitous technologies and the Internet of Things (IoT) finds our brains unprepared for handling the sheer volume of information produced daily by a multitude of sources. Attention deficit disorders, the multi-tasking illusion, learning difficulties, sleep deprivation, weak memory, chronic stress, and others, are just a few examples of the negative side effects that modern technologies impose on everyday life. As modern technologies become increasingly pervasive and even addictive [174], while blurring the line between personal and professional life, such negative side effects are expected to further exacerbate. In response, some companies limit access to technology after a certain hour (e.g., the "*right to disconnect*"), while individuals decide to abstain from using smart devices or social media for a period of time or even completely. But why should one have to resort to such practices after all? Is technology per se not meant to help us improve our quality of life, and eventually realize our full potential as human beings?

Several endeavours promise to bring closer together the human mind with technology, in what has been named "*Human-Machine Confluence*", essential the vision in which the human brain converges with the machine [81]. An EU project under this title has attempted to showcase that the concept may be viable in the future, identifying a set of research challenges 10 years ago in which very few advancements have happened since then. In the US, the *BRAIN* initiative[11] received initial funding of approximately $110 million from the Defence Advanced Research Projects Agency (DARPA), the National Institutes of Health (NIH), and the National Science Foundation (NSF). The EU *Human Brain* project[12], involving researchers from over 100 institutions, received funding over one billion Euros, together with criticism from Europe's leading neuroscientists. More recently, SpaceX and Tesla CEO Elon Musk has joined the BCI venture with a newly founded company called Neuralink[13]. This company is centred on creating implantable interfaces in the human brain, with the eventual purpose of helping human beings merge with software for a true human-machine symbiosis. Facebook unveiled a project on a BCI that could also be used by patients with severe paralysis. This will be a system that allows one to type even faster than with one's physical hands, at upwards of 100 words per minute.

We believe that modern technologies hold the potential to greatly amplify human cognition in the entirety of its spectrum, seizing in a way the role of natural evolution [205]. In this work, we argue that the negative effects observed by increased technological use are simply side effects of our inability to keep up with the pace in which technology advances. We attribute this phenomenon to the fact that despite the overall technological proliferation, the devices and systems (i.e., machine side) we use daily, still remain oblivious of our cognitive and affec-

[11] `https://www.braininitiative.nih.gov`

[12] `https://www.humanbrainproject.eu`

[13] `https://neuralink.com`

tive states, assuming always the maximum of our cognitive capacities. We name this discrepancy "*cognitive gap*", and we theorize that an intermediate software architecture (see Chapter 10), between human and machine could bridge this gap, essentially paving the way towards human-machine convergence [170].

Chapter 3

Augmenting and Measuring Memory Recall

In this chapter, we present the methodology we follow for achieving the aforementioned thesis goals. In principle, we design and develop memory interventions in the form of mobile and desktop applications that are later deployed on participants' devices. Participants are usually volunteers recruited via University newsletters (and advertisements) and agree to go through the entire study design for testing a memory intervention. Each intervention, aims at testing a different aspect of augmenting one's memory recall. Often, participants are also invited to our lab for additional tests. So far, we have conducted (or participated in conducting) a set of 6 deployments with a total of 199 participants, in field studies that lasted from one hour up to 5 weeks. In the following sections, we first present our contextual information sources for generating memory cues, followed by a section on cued recall, the primary psychological theory behind most of the trials for augmenting one's memory with technology. Next, we describe the apparatus (i.e., equipment) we have used for data collection, intervention delivery, and feedback acquisition. Then, we outline our evaluation methodology, where we present the methods we have used for evaluating the effectiveness of our interventions throughout this work. Finally, we present a brief overview of the statistical analyses we typically perform for assessing the effectiveness of the interventions tested in each deployment.

3.1 Memory Cues

Given the great effectiveness of pictures (and video) in aiding memory recall [48], we focus on visual information as the central means for tailoring "*primary*

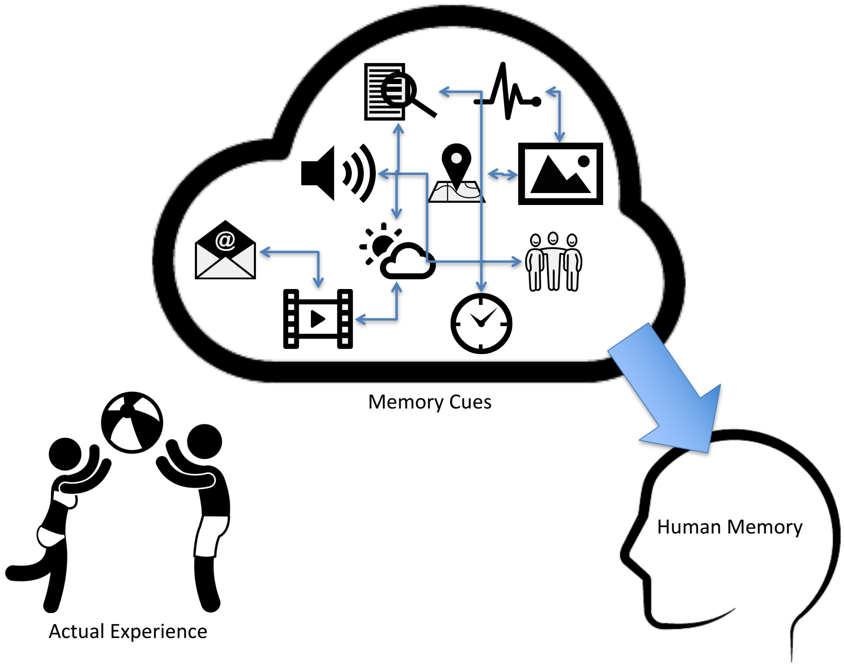


Figure 3.1. Contextual information sources that produce cues for supporting one's ability to recall a past experience or prior knowledge.

memory cues", with video and images comprising the "*primary contextual information sources*". Nevertheless, additional contextual information sources may produce "*secondary memory cues*" that can either enhance visual (i.e., primary) cues or independently be used as memory cues (see Figure 3.1). Typical secondary memory cues are *location*, *time* and *social context* (i.e., co-presence), though each comes with its own strengths and weaknesses. For example, location cues have been found to support recall by eliciting patterns of behaviour (e.g., "*Here I was on my way to work and hence I was feeling stressed*") [125]. However, location cues need to vary significantly (e.g., work and home) for supporting recall without the aid of additional cue information [243]. Time, on the other hand, is central to episodic memory, as episodic memories are clustered together in episodes and placed in a temporal order when being registered in memory [48].

Visual information as captured via pictures and videos often include people's co-presence and interactions in the form of social context. Social context cues, in the form of people we encounter, are often assumed effective memory cues for triggering autobiographical memories. In fact, an analysis of which elements within SenseCam pictures provide the best triggers for memory recall showed that people in pictures were often associated with vivid recollections [144]. In particular, for pictures that contain people that one considers important in one's life (e.g., family members, friends, etc.), personal significance makes a cue more effective in aiding recall, as people tend to pay more attention to personally significant details of an experience. In fact, visual information does not only benefit episodic memory, but semantic memory too, since episodic and semantic memory are closely intertwined [227]. Semantic memory is believed to be a highly organized network of concepts, words and images [231], and thus textual memory cues in the form of words (e.g., category names), have been used extensively for jogging it in lab settings [193]. Thus, we believe visual information in the form of words that convey relevant meaning for one (e.g., the topic of a prior conversation) can be utilized for forming semantic memory cues that target semantic memory.

In addition to the classic primary (visual) and secondary memory cues, one can think of a third set of cues, which we could call *alternative cues*. Alternative cues can be viewed as a subcategory of secondary cues that also stem from additional contextual information. However, due to their atypical nature and lack of research on their ability to cue recall, we differentiate them from established secondary memory cues. Alternative cues can come in many, often unexpected forms: Internet browsing history, prior mobile application use, Wi-Fi SSID names, heart rate and electro-dermal activity (EDA) levels, are only a few examples. Despite their peculiarity, alternative cues may lead to inference of additional cues or help filter primary memory cues, and thus hold potential in helping someone to remember. For example, a Wi-Fi SSID with the name "USI" (i.e., Università della Svizzera italiana) reveals alternative topological information in case actual location is absent or cannot be recalled.

3.2 Cued Recall

As previously mentioned, we use technology as a medium for applying established memory theory and its methods in everyday life settings, with an aim to enhance one's memory. The primary psychological memory intervention method, on which this work is based, is *cued recall*: the practice of augmenting one's memory with memory cues [35]. *Cued recall* is the retrieval of a memory with the support of cues. Such cues can often be semantic, supporting both episodic and semantic memory recall, while involving the activation of the hippocampus [193]. Cued recall differs from *free recall* in that a cue is provided (e.g., a word) which is related to the memory one is attempting to recall, with the aim of enhancing one's ability to recall that memory. Instead, during free recall, no cues

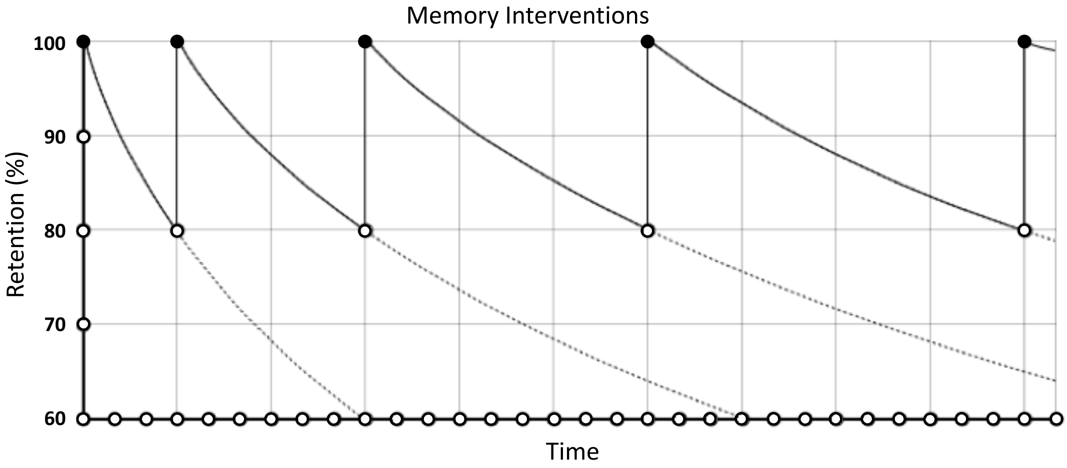


Figure 3.2. The theorized effect of cued recall over time on human memory, also known as "spaced repetition" [9]. After a number of times cued recall has been applied, memory retention decline rate should slow down, turning progressively from exponential to linear.

are provided and one solely relies on one's memory for recalling. Typical cued recall tasks in Psychology are the provision of category names, in which words are originally grouped, and the presentation of related words, for recalling a series of spoken out words. For example, for remembering the word "tree", the word "branch" could be used as a cue in a cued recall task. As a practice, cued recall has been successfully used for supporting patients suffering from Amnesia, Alzheimer, and Dementia [35, 90]. When delivered over certain periods and at certain time intervals (e.g., once a day), cued recall is expected to have a beneficial effect on human memory, improving one's ability to recall a specific past experience or prior knowledge. In fact, we expect that the natural memory retention decline rate due to forgetting will gradually decelerate (see Figure 3.2).

In this work, we adapt the traditional cued recall method to the mobile application context (and not only), while extending it to all types of memory cues, and primarily visual cues (i.e., originating from pictures, videos, etc.) for enhancing the recall of one's personal memories or prior knowledge. However, for successfully triggering one's autobiographical and semantic recollections automatically, the utilization of innately personal and custom-tailored memory cues (e.g., from pictures of one's daily life) is required. Drawing on prior work in the field of human memory augmentation, where lifelogging images significantly improved the episodic memory of patients suffering from memory loss (i.e., SenseCam [111]), implies that our approach could be rather promising. In fact, the novelty of our approach lies in the exploration and evaluation of novel contextual information for the generation of effective memory cues that are displayed throughout one's day for augmenting one's memory recall in diverse contexts. This yields an unprecedented opportunity for eliciting insights that will inform the design and development of future pervasive memory augmentation systems, implementing an end-to-end (from capture to presentation), customized memory augmentation, that will be delivered in situ, unobtrusively, and continuously via ubiquitous technologies. We expect that the memory augmentation effect will persist even when the system seizes to provide further memory support, greatly reducing the rate at which (episodic and semantic) memories are forgotten (see Figure 3.2).

3.3 Apparatus

In this work, we use a diverse set of devices that can be grouped by purpose in 3 broad categories: *data collection*, *intervention delivery*, and *feedback acquisition*. The data collection category includes devices that capture data that is used in the selection and generation of memory cues. The intervention delivery category includes devices that deliver memory cues (e.g., a smartphone), whereas the feedback acquisition category involves devices purposed for obtaining data used for evaluation purposes (e.g., how effective a memory intervention was). Some devices may serve more than one role simultaneously. In general, we use the following devices:

- **Smartphones** are the main instruments for *delivering our memory intervention* in the form of presenting memory cues. Alternative options would be presenting memory cues on domestic or even public ambient displays, smart watches or VR/AR headsets (see next). A smartphone is typically always available and within one's reach. Thus, we can assume a maximisation of our memory intervention by capitalizing on increased user exposure to memory cues displayed on one's smartphone. A smartphone can also have a *feedback acquisition* role for collecting in situ feedback regarding the memory intervention tested each time, participants' well-being, affective state and others, typically via Experience Sampling Method (ESM) [45], Likert scales (e.g., System Usability Scale [32]) and questionnaires. Furthermore, a smartphone can also be used for *data collection* that complements the process of delivering memory cues. For example, continuous location monitoring can be used for displaying different memory cues when one is at home and when at the workplace. We usually deploy our application prototypes on participants' personal devices, hence drawing on the familiarity they have with their own devices. When certain device criteria are not met (e.g., screen size, OS version, etc.) by one's personal mobile device, we provide one with an eligible device available in our lab.

Figure 3.3. Left: Narrative Clip 2, Right: Empatica E4.

- **Narrative Clip**[1] is a compact, modern and wearable lifelogging camera that enables the collection of images automatically and manually, by simply double-tapping it (see Figure 3.3). One can simply attach it to one's clothes and it would capture pictures throughout one's experience (e.g., a visit in a museum). The Narrative Clip can be considered the descendant of Microsoft's SenseCam that eventually made the concept of lifelogging known and somewhat popular. The newer version (Narrative Clip 2), comes with an improved resolution, increased storage capacity and an accompanying mobile application that allows one to customize picture capture interval time, while supporting the capture of short (2 min) video segments. We use this device for collecting pictures for supplying our memory cue selection and generation process. Hence, the Narrative Clip serves as a *data collection* tool.
- **GoPro HERO4**[2] is a small portable camera (41×59×30 mm) equipped with a wide-angle lens that weighs 83 grams (147 grams with case), and can capture up to 4K resolution videos (3840×2160) and 12 MP pictures.

[1] http://getnarrative.com/
[2] https://gopro.com/

Figure 3.4. GoPro HERO4 with an external directional microphone, a tripod, and a Wi-Fi remote controller.

HERO4 supports Bluetooth and Wi-Fi connectivity, and comes with a Wi-Fi remote that can pause and resume recording. HERO4 is particularly popular for capturing 1st-person content during (extreme) sports or other outdoor encounters (e.g., time-lapse video), as it is notoriously durable and water-resistant up to 40 m (with case). We utilized HERO4 primarily for its easy-to-deploy nature and for the high quality of recorded video. For improving the quality of recorded audio, we have used a directional microphone that reduces noises from other sources, while for improving the angle of capture, we have used a small tripod (see Figure 3.4). We typically deploy HERO4 for recording participants utterances during a follow-up interview, but also for collecting video that would be later utilized for generating memory cues (e.g., Chapter 7). Therefore, HERO4 serves as both a *data collection* and *feedback acquisition* tool.

- **Empatica E4**[3] is a wristband equipped with a photoplethysmography (PPG) sensor for collecting physiological feedback of close to clinical standards accuracy, while remaining unobtrusive to the engagement one has with an experience (see Figure 3.3). It monitors a series of physiological responses such as Electrodermal Activity (EDA/GSR), activity levels (with 3-axis accelerometers), body temperature, Heart Rate (HR), and Inter-Beat Interval (IBI) as a measure of Heart Rate Variability (HRV). Physiological responses can be used for estimating one's arousal and partially one's affective state in situ, and is thus a useful indicator of one's engagement in an experience. Due to indications that physiological responses help us refine our memory cues (e.g., help filter out irrelevant information or indicate the effect of a memory cue), we consider the E4 wristband both a *feedback acquisition* and *data collection* tool.
- **Smart watches** (and fitness trackers) are one of the latest "consumer trends" in the field of wearable computing. Smart watches are bigger than smart wristbands (e.g., Fitbit[4] and Jawbone[5]) because they are also equipped with a touch screen at the size and shape of typical wristwatches. Smart watches can host a large variety of sensors, some of which are dedicated for monitoring physiological responses. For example, an optical sensor (i.e., PPG) embedded in the back-side that touches the inner part of one's wrist can detect one's current heart rate. Though yet not able of achieving the accuracy of professional medical instruments, or that of E4 wristband, it can however provide a clue of one's current physiological state (e.g., arousal levels). Thus, smart watches and fitness trackers can be used for both feedback acquisition and data collection purposes. Furthermore, a smart watch is designed to function as the extension of smartphones, delivering all types of notifications (e.g., SMS, calls, e-mails, etc.). Despite being equipped with a small screen, we believe that the smart watch holds a potential for displaying memory cues, and hence perhaps able to deliver a memory augmentation (e.g., display a grocery icon for reminding one to do the groceries). We thus consider the smart watch also capable of delivering a *memory intervention*, playing the same (triple) role as a smartphone.
- **Tobii TX300**[6] stationary eye tracker is the primary instrument we use in this work for analysing how one's eye gaze is changing when viewing different memory cues on novel (mobile) Lifelogging User Interfaces (LUIs). For

[3] https://store.empatica.com/products/e4-wristband?variant=945527715/
[4] https://www.fitbit.com/eu/home/
[5] https://jawbone.com/
[6] http://www.tobiipro.com/product-listing/tobii-pro-tx300/

this, we typically deploy the mobile application prototype we would like to investigate on a mobile emulator software running on the eye tracker desktop computer. Participants are then asked to perform some tasks using our emulated mobile application prototype on the eye tracker while we are measuring their eye gaze patterns. After the investigated areas have been divided into AOIs (Areas of Interest), specialized software encodes recorded eye gaze patterns into fixations, saccades, blinks and other eye gaze descriptive variables. For example, fixations indicate the number of times eye gaze focused on an AOI, while saccades the number of subtle movements between fixations. Typically, these measures are used for eliciting what areas (e.g., which memory cues) draw more attention and infer the logic behind decisions made when using an interface (e.g., looking at a picture before clicking a LUI button). We use the eye tracker solely for *feedback acquisition* purposes.

- **Oculus Rift**[7] is a Virtual Reality (VR) head mounted display (HMD) that is used for completely immersing a user inside VR experiences. Oculus features an organic light-emitting diode (OLED) display with a resolution of 1080x1200 per eye, a refresh rate of 90 Hz, and a 110° field of view. It is also equipped with integrated headphones, which provide a 3D audio effect, rotational and positional tracking. The positional tracking system utilizes a USB stationary infrared (IR) sensor that receives light emitted by IR light-emitting diodes (LEDs) integrated into the HMD. The sensor is typically located in front of the user (e.g., on a desk), generating a 3D space, that enables the user to use the Oculus while standing, sitting or walking around the room. In this work, we have used Oculus Rift as a VR experience presentation modality, for detecting among others which characteristics in a VR scene (e.g., perspective, scene transitions, etc.) lead to vivid VR recollections (see Chapter 8). Thus, we loosely consider Oculus Rift as a memory intervention delivery tool.

3.4 Evaluation in Human-Computer Interaction

Human-Computer Interaction (HCI) is a multi-disciplinary field, situated on the intersection of Computer Science, Cognitive Psychology, Sociology, Human Factors, Artificial Intelligence and Design Engineering. HCI, as a research field, concentrates on **how humans interact with computers via** the use of **interfaces**. HCI strives for understanding how humans use technology, introduces tools and

[7] https://www.oculus.com/rift/

techniques for developing better (usable) systems, and enables *safe*, *efficient*, and *effective* interaction. If one tries to summarize HCI, it would boil down to the processes of **design, implementation, and evaluation** of interactive systems in the context of the user (e.g., task at hand) [62]. Evaluation is a critical process in HCI, as it tests whether the prototype of a system meets the user requirements, and generally behaves as expected. Evaluation addresses three main aims: (i) to assess the functionality and accessibility of a system prototype, (ii) to assess user experience (UX) during the interaction with the system prototype, and (iii) to detect any flaws in the system prototype. Ideally, evaluation is an iterative process that should run throughout the entire design life cycle, further informing the design of a system prototype, until it satisfies a range of requirements. The iterative design-evaluation process ensures that any flaws detected, will be more effectively addressed earlier in the design phase, than later in the implementation phase, where resources have already been allocated. Typically, evaluation techniques are divided in two broad categories: *expert analysis* and *user studies*. In this work, we have exhaustively tested our prototypes in *user studies*, in mixed in-lab and in-the-wild settings.

3.4.1 User Studies

The aim of user studies is to test system prototypes in pragmatic settings, for example by recruiting real users as participants that try out a system prototype, while undergoing a well-defined study design [143]. Our participant selection process is not sophisticated, since our system prototypes are intended for testing a memory intervention by delivering it to the general audience. Thus, we typically recruit participants by "snowball" sampling, from our University premises, online with deploying a mobile application prototype on a digital application marketplace (e.g., Google Play[8] and the AppStore[9]), or in situ in the context of the study. For example, in Chapter 8, we recruited our participants in the Aquarium of Genova, Italy for assessing the memorability of an underwater VR experience. In fact, we have found that context in terms of locality plays an important role when it comes to experiences that utilize location-aware narratives [128]. Albeit we opt in for testing our prototypes in the wild for unveiling plausible effects on our participants' behaviours, it is often the case that a study design may require in-lab testing. This normally happens when the use of special equipment is involved, such as an eye-tracker (e.g., Chapter 5), or when it comes to

[8] https://play.google.com/
[9] https://www.appstore.com/

applying a specialized technique, such as memory testing (e.g., Chapter 6). The number of participants is usually defined by taking into account the undergoing study design, while utilizing classic techniques outlined by Cook and Campbell in 1979 for establishing statistically significant results [50]. Often, before testing a system prototype with a full set of participants, we run limited "focus groups" or "pilot testing" sessions for ensuring that the prototype functions as intended and the study design is solid. In both, pilot and full-scale user studies, we adhere to established, ethically-sound practices, as outlined in the respective Ethics section in Chapter 1, and we strive for maintaining high standards of external (ecological) and internal validity.

3.4.2 Quantitative and Qualitative Methods

Throughout the completion of this work, we have applied both quantitative and qualitative evaluation methods. Quantitative methods amass numeric data, such as user performance metrics or opinion ratings [101]. Instead, qualitative methods aggregate non-numeric and descriptive data, usually illustrating the level of UX with the system prototype, or a specific problem during use. In our rigorous study designs, we have utilized both quantitative and qualitative methods in conjunction, for explaining the observed phenomena. In the following sections, we provide an overview of the methods we have used.

Quantitative

The majority of our quantitative methods utilize self-reporting measures, typically in the form of Likert-scales. Despite self-reporting methods are highly subjective, thus often called "subjective measures", they produce numerical data, and hence are considered quantitative. Nevertheless, we also employ "innately" numerical quantitative methods, such as physiological responses monitoring and usage logging.

- **Experience Sampling Method (ESM).** It investigates what people do, feel, and think in the settings of daily life. In its original form, ESM asks one to provide systematic self-reports at random occasions throughout a normal week, and is administered in paper form, same as typical questionnaires [142]. Collections of these self-reports can answer questions such as, "*How do people spend their time?*", "*How do people feel when engaging in sports?*", or "*What people think when at home?*", and others. Recently, the increasing emphasis on how mobile technologies are experienced in everyday life has resulted in an increased interest in in-situ measurement and, in particular, ESM. Consolvo and Walker have proposed a systematic approach on how to *schedule*, *notify*, and finally *deliver* questions using ESM digitally on smartphones for evaluating ubiquitous computing technologies [46]. ESM is often considered as the "*gold standard*" of in-situ measurement, as it samples experiences and behaviours right at the moment of their occurrence, thus reducing memory and social biases in self-reporting [54, 124]. However, ESM also entails significant drawbacks, such as disrupting a users' current activity and imposing an additional reporting burden [208]. In this work, we have employed ESM in various user studies, in conjunction with Likert-scales for quick numerical entry (e.g., from 1 to 5), primarily for collecting ground truth (e.g., Chapter 5).

- **Day Reconstruction Method (DRM).** Motivated by the aforementioned ESM drawbacks, Kahneman and colleagues proposed the Day Reconstruction Method (DRM) [124], a retrospective, self-report protocol that aims at increasing users' accuracy in reconstructing their experiences at the end of a day. It does so by imposing a chronological order in reconstruction, thus providing a temporal context for the recall of each experience. At the end of the day, participants are asked to recall their day from the beginning in the form of temporally ordered episodes (e.g., waking up, or going to work), while prompted to self-report on thoughts, feelings, and emotions [216]. DRM has been found to provide a reasonably good approximation to ESM, reducing reporting burden and memory biases, while the method has been well adopted also in the HCI community [58]. In this work, we do not utilize the DRM method as-is, but we rather leverage on the DRM concept for augmenting memory recall in the form presenting temporally ordered memory cues that characterize one's daily life episodes.
- **System Usability Scale (SUS)**. It is a simple ten-item questionnaire that provides a general view of subjective assessments of usability of a system or interface [32]. A SUS questionnaire is usually administered right after a system, interface, or a prototype has been used. In the original SUS questionnaire, participants are asked to estimate the extend to which they agree with general statements about the system (e.g., "I found this system unnecessarily complex") in a Likert-scale from 1 ("strongly disagree") to 5 ("strongly agree"). Typically, a system that achieves a usability score of above 68 is considered of above average usability. Due to its prevalence for assessing usability, SUS has been slightly modified to refer to products instead of systems. It has been used for assessing general usability and customer satisfaction but also for comparing different tasks within the

same GUI, different versions of the same system, and competing implementations of a system, while allowing for a competitive assessment of comparable GUIs and different interface technologies [17]. In our work, we use SUS for evaluating the usability of our prototypes in amplifying the memory of our participants.

- **NASA-Task Load Index (NASA-TLX)**. TLX, in short, is a widely used multi-dimensional tool for subjective self-reporting of mental and physical workload when performing a task [100]. TLX is comprised of six item-scales that measure somewhat independent groups of variables: *mental*, *physical*, and *temporal* demands, on one hand — *frustration*, *effort*, and *performance*, on the other hand [99]. TLX posits that the combination of these six variables provides a good approximation of the exhibited workload when performing a task. It is typically administered during the execution of a task or immediately afterwards. Initially, TLX was developed by NASA for assessing the cognitive and physical effort of system operators in the field of aviation. For decreasing between-rater variability and increasing overall sensitivity, a weighting scheme is often applied: In the beginning of the study, participants are asked to indicate the degree to which they think the aforementioned six variables are pertinent to their definition of workload, by forming a variable ranking through a series of pairwise comparisons. When the TLX is finally administered, each scale rating is multiplied by the appropriate weight, thus providing a customized approach to individual workload definitions in a percentage scale (%) [99]. In this work, we have used the unweighed version of the TLX for assessing the workload entailed by our memory interventions.
- **Affect Grid.** It is a subjective, self-reporting, single-item scale, developed for quickly assessing affect based on two dimensions: arousal (high–low) and valence (positive–negative) [191]. The two dimensions are combined for forming a 9x9 grid, with valence situated on the horizontal axis while arousal is placed on the vertical axis. As such, the center of the grid represents a neutral, average, every day feeling, while being neither positive nor negative. After participants have been instructed how to use it, Affect Grid is a reliable method for eliciting self-related mood, suitable for rapid, sequential assessments in large numbers. In this work, we have used Affect Grid for inferring participants' emotions, with and without the presence of our memory interventions (e.g., Chapter 7).
- **Presence Questionnaire (PQ)**. Presence is defined as the subjective experience of being in one place or environment, even when one is situated in another, and it is a particularly relevant measure when assessing the

effectiveness of virtual environments and VR experiences [247]. The original PQ (version 1) encompasses a set of 32 seven-point Likert scale items, clustered into six subscales, namely: sensory exploration, involvement, interface awareness, control responsiveness, reality/fidelity and adjustment/adaptation [246]. In general, PQ has been found to positively relate with measures of task performance (e.g., NASA-TLX). It is often the case, that an experiment does not involve the entirety of items enlisted in PQ (e.g., when no haptics are involved), and thus it can be modified to best fit the user study at hand. In this work, we have utilized PQ for measuring the degree of presence in VR experiences, as an element that facilitates memory recall (see Chapter 8).

- **Memory Testing.** It involves a group of methods applied in conjuction or in a follow-up fashion rather than a distinct method per se. In several user studies, we employ diverse memory testing methods for assessing a baseline and/or the effect of our memory interventions. Typically, we first provide our participants with a "target experience" to be remembered, such as a campus tour (e.g., Chapter 6), a VR tour (e.g., Chapter 8), or simply pick some of their daily life experiences (e.g., Chapters 5 & 7). Then, we use **free recall** tests for assessing how much one is able to recall about the "target experience" and/or the "prior knowledge", when solely relying on one's memory. In a free recall test, participants are asked to recall a past target experience or prior knowledge, while instructed to cluster their memories in some form of episodes (e.g., by location or time), and subsequently write down or report out loud their recollections. Since the notion of temporal order is highly pervasive in episodic memory, participants in free recall, exhibit the *primacy* and *recency* effects, according which they are particularly able to recall the first and the last episodes of a past experience, respectively, with worse performance for those episodes in between [162]. A variation of free recall is the **delayed recall** test, where participants are again asked to recall a past target experience and/or prior knowledge, but after a certain time period has elapsed. A delayed recall test is applied in conjunction with a prior free recall test, as a measure of memory deterioration. In our user studies, we use a standard interval of one week (since the target experience has occurred and/or the target knowledge has been acquired) for a delayed recall task, as this interval entails an approximate 75–78 % natural memory retention decline over the target experience, when no rehearsals have been performed in the meantime (see Ebbinghaus' forgetting curve Figure 2.1). Finally, a **cued recall** test [35, 90], as previously described, aims at assessing participants' ability to recall an experience and/or prior knowledge after delivering our memory intervention in the form of memory cues. Recollections originating from different memory test modalities are then coded and rated by multiple, independent coders and raters, that compare participants' recollections with the full record of the target experience and/or knowledge, usually kept in a written form (e.g., the transcripts of a meeting). The reliability of the independently produced scores is meticulously tested, by applying inter-coder and inter-rater reliability methods (see relevant section in Statistics), minimizing biases and discrepancies. Independent recall scores are then averaged for producing overall recall scores in a percentage scale (%). Then, we compute the difference (Δ) between different recall scores (e.g., delayed recall and cued recall) for assessing the effectiveness of our employed memory intervention.

- **Physiological monitoring.** Physiological responses have recently become increasingly available, due to an uptake to physiological monitoring equipment (e.g., FitBit), and can be indicators of intrinsic states such as emotions/feelings, arousal levels, affective and cognitive states (e.g., attention). Thus, physiological responses, if properly analysed, can provide useful insights about the UX of a prototype [239], hint at the effectiveness of a memory intervention [134], help in the cue generation process [168, 200], or even indicate opportune moments for delivering memory cues [95, 182]. In fact, physiological responses such as, Heart Rate Variability (HRV), brainwave activity (via EEG), and blink interval (via eye-tracking), can indicate cognitive workload [95, 245], and have been found to significantly correlate with prominent self-reporting workload methods such as NASA-TLX [196, 254]. In this work, we utilize 2 categories of physiological responses, (i) bodily activity, and (ii) eye-gaze activity:

  (i) **Bodily activity:** HRV (variability in the time between heartbeats), Blood Volume Pulse (BVP) in millivolts (mV), Electro-dermal activity (EDA) in microsiemens (μS), skin temperature in Celsius (C°) and 3D–axis acceleration in m/sec^2.

  (ii) **Eye-Gaze activity:** Visit count and visit duration (i.e., number of times and duration of eye-gaze in a certain Area of Interest – AOI).

- **Usage logging.** This primarily quantitative method is used in the case a memory intervention is deployed as a mobile application on participants' smartphones. Several usage metrics are collected for gaining additional insights on how participants used our prototypes. These metrics may describe specific usage behaviour with a deployed prototype, such as number of times the prototype has been launched, the time a memory cue was viewed, overall duration a memory cue was viewed, subtle bar cursor movements, button touches, and others. Additionally, we may log more generic usage behaviour, such as overall mobile application usage (e.g., application usage), for understanding how our memory interventions were used in conjunction with other mobile applications.

Qualitative

Despite the strong reliance of science on quantitative and experimental methods, in a interdisciplinary field such as HCI, it is often the case that complex, social phenomena cannot be easily quantified or experimentally manipulated, or even ethically investigated with experiments [37]. A typical example of a concept that cannot be investigated with quantitative methods is Privacy, as it cannot be reduced to just numbers. Hence, the primary focus of qualitative methods is to investigate the *why* and *how* of the decision-making process, and shed ample light on an observed behaviour. In this work, we utilize a specific set of qualitative methods, in combination with the aforementioned quantitative methods for gaining additional insights over the observed phenomena. We briefly describe the qualitative methods we use below:

- **Retrospective Think-Aloud (RTA).** The RTA is a usability evaluation method that collects a participant's utterances after a task is completed [92]. It differs from the most frequently used approach, the Think-aloud (aka. Concurrent Think-aloud), in which participants are actively verbalizing thoughts and actions while performing a task [72]. Interestingly, the most widely known form of RTA, the "stimulated" RTA, is directly aligned with the essence of this work, since it asks participants to report on thoughts and actions retrospectively, using collected stimuli as memory cues. For example, in stimulated RTA, a participant is asked to perform a task on the eye-tracker, while a dual-view video containing the participant's actions and facial expressions is captured (see 5). When later asked to review the video and Think-aloud, the participant has increased provision of memory cues for vividly recalling and accurately describing thoughts, actions and the general decision-making process. Stimulated RTA is considered to alleviate the participant from the negative effect on task performance, due to the act of speaking while performing an actual task. In fact, verbalizing information while performing a task may affect a participant's attention and focus [92].

- **Cognitive Interview.** The cognitive interview is an interviewing technique used by crime investigators for interviewing eyewitnesses [78]. In a cognitive interview, the interviewer constantly prompts the interviewee to recall more and more details (e.g., by prompting one: "*And then what happened?*") while taking notes. Once the interviewee is not able to recall any more details, the interviewer will go through the notes and read out different parts to extract additional details (i.e., "*Earlier you said... Can you tell me a bit more about that?*"). In our work, this method has not only served for improving the level of detail recalled by the participant, but also helps minimize misinterpretations. We typically perform the cognitive interviews in a quiet office, while recording them using a voice recorder. Recorded utterances are then send to a trusted 3rd party service for transcription. Finally, transcripts are then either automatically processed, utilizing text processing algorithms and tools (e.g., the LDA algorithm [25]), or manually coded and rated, as mentioned before (see Chapter 7).
- **Open-ended Interview.** Interviewing, as a method, has been pivotal for the field of social sciences, and has been traditionally used for eliciting reports of attitudes and perceptions [187]. In HCI, an open-ended interview is a way of gathering information by asking participants to answer questions. Questions can be scripted, but when the interviewer does not anticipate a specific type of response, the interview is considered open-ended. An open-ended interview may involve the collection of factual data such as demographics, but its aim is to focus on participant's thoughts, feelings, experiences, knowledge, skills, ideas, and preferences. In this work, we perform an open-ended interview at the end of trials, when participants have already used our memory intervention, (typically after TLX and SUS, if applied), inquiring into participants' impressions, and overall experience. So far, this method has been found particularly effective in detecting usability flaws, and thus helped us to improve our prototypes in the next iterations. Participants' utterances are usually recorded and the transcribed. Transcripts can later be subject to thematic analysis, for identifying, analysing, and reporting patterns within data [30]. Typically, we apply open-coding for inferring concepts, categories, and properties, using specialized software (e.g., Atlas.ti[10]) [37].

[10] `http://atlasti.com`

3.4.3 Statistics

After data has been collected, statistical tests and analyses are performed to elicit the effectiveness of the intervention tested in each deployment, and thus help us answer previously stated research questions, or even inform our prototype design. The statistical analyses performed depend on the data and the purpose of the experiment. Nevertheless, in the case of human subject research statistics, we follow standard statistical evaluation procedures, as described in Field's excellent textbook on "Discovering statistics using IBM SPSS statistics" [77]. First, we assess the normality of distribution for collected data using Kolmogorov-Smirnov or Shapiro-Wilk normality tests, excluding outliers and extreme values. Levene's tests of homogeneity of variance are performed for investigating if our data is homogeneous. Depending on our study design (within-between subjects) and the levels of our independent variable, we apply a range of parametric tests such as, t-tests, chi-square, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA), and others. Additional pre-tests may be required such as Sphericity tests for repeated measures ANOVA. In case our data is not normal, we apply non parametric tests such as, Friedman test, Kruskal-Wallis test, Mann-Whitney U test, Mood's Median test, and others. Significant effects are always reported at a p-value of less than 5 % (i.e., $p < .05$). Post hoc tests using the Bonferroni correction are used for uncovering differences between the levels of the independent variable. When data is not normal, Wilcoxon Signed-Rank tests are used instead. Correlations are investigated using Pearson for normal data and Spearman for non normal data (i.e., ranks and scales). These analyses are performed on SPSS version 22 statistical package by IBM.

Testing Reliability

In the case of manual coding and rating by independent researchers, we assess inter-coder reliability with a variant of Krippendorf's Alpha inter-coder reliability test, known as multi-valued coding [136]. Typically, a topic subset size of 10 % is enough for assessing inter-coder reliability [38], and an alpha value above 80 % indicates good inter-coder reliability [241]. For assessing inter-rater reliability, we simply calculate the percentage of overlapping topic matches.

Dealing with Physiological Data

In the case of physiological data, further processing is necessary. Usually, different physiological data streams have to be synchronized before analysed, through custom Python scripts. Typical signal processing techniques are applied for removing noise and outliers, before peaks or other features can be computed. For this, we use either Python or Matlab. However, not everyone exhibits the same physiological reaction when exposed to the same experience. This is mainly due to interpersonal differences, such as different health, mental, and affective states, resulting in largely divergent physiological measurements. For example, a normal resting heart rate for adults ranges from 60 to 100 bpm. For dealing with such interpersonal differences, we usually compute an individual physiological baseline for each participant, while subjected to no intervention yet (i.e., at resting state), and compute the difference Δ with the physiological responses captured during intervention delivery. An alternative route is the z–transform method, allowing us to compare samples with diverse distributions. The z–transform method, also called standardization or auto-scalling, computes z–scores by deducing from each value in the sample (x_i) the mean value of the entire sample ($\bar{x}$), and dividing the remainder by the standard deviation of the sample (S) (see equation 3.1).

$$z_i = \frac{x_i - \bar{x}}{S} \tag{3.1}$$

3.5 Summary

In this Chapter, we formally introduced "cued recall", the core psychological method on which this work draws for effectively augmenting human memory recall, and we presented our primary and secondary contextual information sources for generating effective memory cues. We also described how we evaluated our interventions by implementing study designs that have lasted even longer than a month, in both lab and everyday–life settings. Then, we demonstrated the apparatus we have used throughout the work reported in this thesis, and exhaustively reported the methods (both quantitative and qualitative) we have applied so far. In the last part, we provided an overview of the statistical tests and analyses we have carried out in this work. We underscored how we ensure high standards of internal validity in the course of this work by discussing statistical methods for testing reliability and comparing data with wildly varying distributions (e.g., physiological data).

The next part of this thesis focuses on showcasing tangible memory augmentation through a series of field studies and lab experiments. In particular, the next chapter describes a field deployment that tested the effectiveness and viability of mobile cue-based memory augmentation approach in the field as a proof of concept, by utilizing memory cues produced via event-driven capture.

Part II

Studies

Chapter 4

Event-Driven Image Capture for Augmenting Episodic Memory Recall

In this Chapter, we introduce EmoSnaps, a mobile application that captures unobtrusively and automatically pictures of one's facial expressions throughout the day, and uses them as **memory cues** for supporting later episodic memory recall, and inferring one's momentary emotions. We describe a field study that employs EmoSnaps in an attempt to investigate *if* and *how* individuals and their relevant others infer emotions from self-face and familiar face pictures, respectively, as a proof of concept. The study contrasted users' recalled emotions as inferred from EmoSnaps' self-face pictures (i.e., selfies) to ground truth data as derived from Experience Sampling Method (ESM). Contrary to our expectations, we found that people are better able to infer their past emotions from a self-face picture the longer the time has elapsed since capture, a useful insight for the design and development of future pervasive memory augmentation systems. All in all, the findings reported in this chapter provide evidence for the viability and effectiveness of event-driven mobile capture in producing memory cues that substantially augment episodic memory in the field (i.e., RQ1), or else what we call "*(mobile) cue-based augmented memory recall*".

4.1 Author's Contribution

The author of this thesis had a leading role in the study reported in this chapter. In particular, his contribution includes the conceptualization, development and deployment of the memory intervention, the study design, and the data analyses conducted in the original publication [165]. The co-author of the original publication provided his insightful guidance and knowledge on data analysis, while editing sections in the original publication. For more information, see the original publication [165].

4.2 Introduction

Emotions are so tightly connected to facial expressions that one could even question whether there can be emotion without facial expression [69]. Not only it is difficult for people to hide their emotions in facial expressions, research has also shown that humans are surprisingly accurate in recognizing basic emotions, such as anger, disgust, fear, joy, sadness, and surprise, from facial expressions [173]. In particular, when it comes to happiness, research has revealed that humans can accurately recognize the emotion from facial expressions in 96.4 % and 89.2 % of the times in Western and non-Western cultures, respectively [192]. Algorithmic techniques in emotion recognition have flourished [10, 42, 176] and provide a promising approach in stationary settings. On the contrary, mobile settings introduce substantial complications in capturing facial expressions. Some novel solutions have been proposed by Teeters, Kaliouby, and Picard on "Self-Cam" [222], a chest-mounted camera that is able to detect 24 feature points on the face and extract emotions using dynamic Bayesian Models, as well as Gruebler and Suzuki [91] on a wearable interface device that can detect facial bioelectrical signals. While providing the ability to capture emotions in a continuous fashion, both these approaches are highly intrusive, inducing a feeling of being monitored as well as raise concerns of social acceptance, especially when long-term deployments in real-life settings are concerned. With EmoSnaps, we aimed at creating a tool that can be truly transparent in daily life and can be employed in long-term field studies.

In our line of research, we attempt to contribute towards a next step in the field of pervasive human memory augmentation [56]. In particular, we utilize mobile sensor technology for passively logging user context throughout the day, and we utilize collected data for creating memory cues that assist the recollection of one's daily activities and experiences. This work introduces EmoSnaps, a mobile application that captures unobtrusively pictures of one's facial expressions throughout the day and uses them for the later recall of one's momentary emotions. In the remainder of this chapter, we first briefly present the theoretical basis on emotion and memory, followed by an elaboration on the EmoSnaps solution. Next, we describe a two-week-long deployment of EmoSnaps that inquired into *if* and *how* self-face pictures assist the reconstruction of momentary emotions. We then discuss our results and we conclude.

4.3 Memories and Emotions

For quite long, it was believed that memory functions as a "storehouse of past impressions", where experiences are stored and retrieved on demand. The first to question this simplistic approach to memory was Bartlett, who suggested that remembering is rather an act of reconstruction than an act of reproduction [20]. Bartlett claimed that a past event cannot be stored in memory and reproduced as it actually took place, but instead, when recalling, memory provides a representation of a past event that is often distorted. As we have seen, Tulving was the first to introduce the concept of episodic memory as a system that receives and stores information about temporary episodes or events, while concurrently mapping temporal and spatial relationships among them [226]. Based on Tulving's distinction of episodic and semantic memory, Robinson and Clore [189] proposed an accessibility model of emotion, according which "*an emotional experience can neither be stored nor retrieved*", but it can be inferred from contextual details residing in episodic memory. In other words, Robinson and Clore's model assumes that when we recall how we felt during a past event, we first reconstruct the happenings of the event, and then we infer our emotions based on how we think we would feel in those circumstances. Therefore, it is expected that an improvement in one's ability to recall contextual details from episodic memory will help to increase the validity of retrospective self-reporting on experience.

As we have previously seen, a sizeable body of research is dedicated to how we can improve one's recall from episodic memory. According to Tulving, episodic memory hosts contextual information regarding who, what, where, and when [226]. Thus, remembering can be supported by external cues, such as co-presence (social context), visual and audible cues (e.g., pictures, video, or sound), location, and time [19]. Social interactions have been proven to be one of the most effective cues for triggering autobiographic (i.e., episodic) memories. For example, Lee and Dey [145] used SenseCam [111] pictures to investigate what elements included in a picture can enhance memory recall and found that the co-presence of people in images was often associated with rich recollections. However, social context derived from mobile communications data (e.g., SMS) was found to be less effective in assisting episodic recall, mainly due to lack of novelty as thought by the participants [85]. Another interesting approach in supporting episodic memory recall is presented in "AffectCam" [200]. Bodily arousal, as measured via galvanic skin response sensors, is used to distinguish among SenseCam pictures captured during higher and lower arousal. Pictures of higher arousal were found to support richer episodic memory recall when compared to those of lower arousal.

In Chapter 2, we described the primarily visual nature of episodic memory, and thus visual cues have been proven to be exceptionally effective in assisting the recall process [226]. Other memory cues, such as location, lack the immediateness that visual cues induce during recall, and they have been found to implicitly support remembering through enabling inferences from established patterns of behaviour rather than a true recollection of an event [125]. However, location cues need to vary significantly to single-handedly support episodic recall [243]. Time also plays a major role in recall since it is the main driver according to which personal events are registered in episodic memory [226]. Temporal cues have been prevalently employed in retrospective interviews, where recalling the specific time of the day when a particular event happened also assists recalling temporarily adjacent events [85]. A more detailed description, available in Chapter 3, summarizes the effectiveness of the aforementioned cues to trigger episodic and autobiographic memories. The current work examines the potential of self-face pictures to serve as **memory cues** for facilitating emotion recall from episodic memory.

4.4 System

EmoSnaps is a mobile application that captures pictures of one's face using the front–facing camera of one's mobile device (see Figure 4.1). EmoSnaps employs event-driven capture, where predefined events, such as "screen unlock", "phone call answer", "SMS sent", and application launches, trigger a picture capture. In this study, we limited the event-driven capture only to "screen unlock" events, as this best ensured proper taking of the user's face: During and immediately after a screen unlock, the mobile device is typically positioned in front of the face, hence making it more likely to actually capture the user's face via the front–facing camera. EmoSnaps is able to capture a picture within 300–500 ms, adding only minimal interaction delay in the process, and hence being almost transparent to the user. Following a successful picture capture, no further capture occurs for the next 5 min.

4.5 Recall or Recognition?

We conducted a two-week-long deployment of EmoSnaps with a total of 14 participants to inquire *if* and *how* self-face pictures assist the reconstruction of momentary emotions by attempting to answer the following research questions:

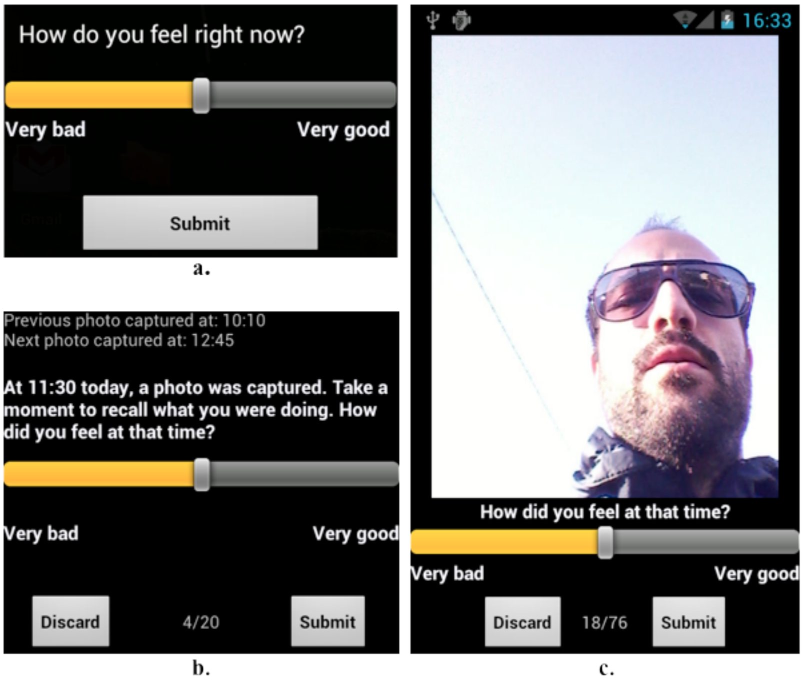


Figure 4.1. a. ESM was used throughout the day to self-report on momentary psychological well-being. b. Time-based reconstruction was employed as a control condition. c. A self-face picture was provided to assist in the reconstruction of momentary emotions.

(a) **Are participants able to recognize their emotions on self-face pictures captured during mobile device usage?** Given prior literature [173, 192], one would expect participants to be able to accurately recognize their own emotions given a self-face picture. However, less accurate emotion recognition could be expected due to the mobile setting, as pictures may be of varying orientation, luminosity, and image quality.

(b) **If participants can accurately elicit their emotions from their self-face pictures, the question is how do they do so?** We can think of at least two ways (Figure 4.2). The first assumes that individuals will recognize their emotions from their facial expressions in their self-face pictures [10, 192]. In contrast, the second option assumes that individuals will use the pictures as **memory cues** for recalling episodic memories (e.g., where they were, what they were doing, who they were with) and, based on this information, will be able to recall their emotions at that given time. Recent work has suggested that emotional experience "*can neither be stored nor retrieved*", but can only be reconstructed on the basis of recalled contextual information from

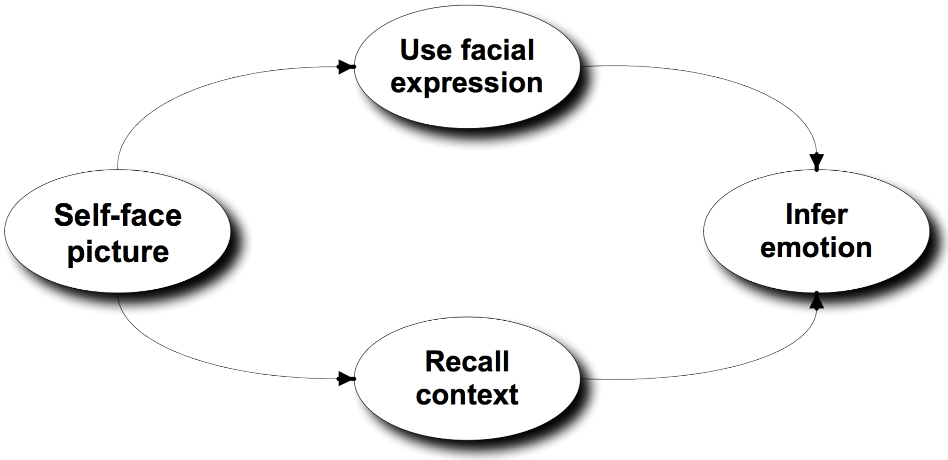


Figure 4.2. If people accurately recognize their emotions, how do they do so? Inferring emotions directly from facial expressions, or recalling episodic memories and drawing upon this knowledge to infer emotion?

episodic memory [189]. If self-face pictures are recent and contain information that may cue episodic memories, participants could as well infer their emotions from these episodic memories rather than infer them from facial expressions displayed in the picture. One could expect these reconstructed emotions to be more accurate than the recognized emotions, given that participants may draw upon rich episodic information in the case of a recent event.

(c) **Are relevant others of individuals better able to infer the individuals' emotions from their facial expressions given the increased exposure to them?** Indeed, research has shown that when participants are subjected to a task of identity matching, reaction times to familiar faces are faster than reaction to unfamiliar faces. Yet, there is no difference in reaction time between familiar and unfamiliar faces in tasks of facial expression matching [253]. Others have shown, however, that familiarity may make a difference by improving the accuracy in recognizing emotions [130]. Given this evidence, we expect familiar others to be better able to infer one's emotions from one's facial expressions.

4.6 Study Design

To address these three research questions, we formed four conditions, each representing a distinct reconstruction process as follows.

- **Photo-Day Reconstruction:** At the end of each day, participants are asked to revisit all self-face pictures taken throughout the day and recall how they were feeling at the time of each captured picture (Figure 4.1c). Pictures are presented in chronological sequence as this has been proven to enhance the reconstruction of episodic memories using memory cues [3]. Thus, we assume participants in this condition to have access to both approaches of emotion inference: *recognition* and *reconstruction*.
- **Time-Day Reconstruction:** At the end of each day, participants are asked to recall what they were doing at the time when a self-face picture was captured, and decide how they were feeling at that time (Figure 4.1b). No actual pictures are shown but are instead stored for Photo-Week condition following next. Instead, the timestamps of the preceding and the succeeding captured self-face pictures are shown, as it might provide a temporal context and thus, assist the reconstruction process [124]. As in the Photo-Day condition, all information is presented in chronological sequence. This type of reconstruction serves as a control condition, and any difference between this and Photo-Day reconstruction in terms of participants' accuracy will be attributed to the effect of the self-face picture.
- **Photo-Week Reconstruction:** A week after the last day of the study, participants are asked to review the total of self-face pictures taken in the Time-Day condition and decide how they were feeling at the time when each picture was captured. They use the same LUI as in the Photo-Day condition (Figure 4.1c), but this condition differs in two respects. First, as a week or more has elapsed since these pictures were taken, we assume participants will be unable to reconstruct episodic memories related to the picture. Second, pictures are presented in random order in an effort to minimize any effect of building contextual knowledge as participants go through the pictures. Thus, in this condition, we assume participants to infer their emotions only from facial expressions.
- **Photo-Relevant Reconstruction:** For each participant, a relevant other is chosen to evaluate the same pictures the participant has evaluated during the Photo-Week reconstruction. Relevant others consisted either of the partners-in-life or the closest colleague and/or friend of each participant. We judged that these groups would have an increased familiarity with par-

ticipants' facial expressions. Relevant others used the same LUI as in Photo-Day and Photo-Week reconstructions (Figure 4.1c) with the pictures being displayed in random order.

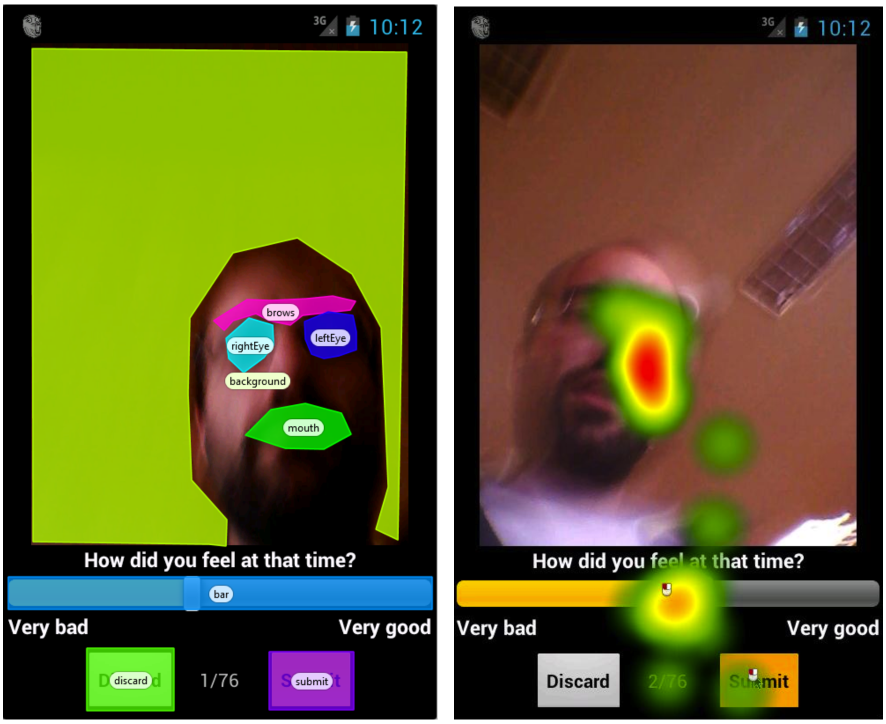
We also formed the following hypotheses:

(H1) In general, participants should be able to infer their emotions from their facial expressions. Albeit we have found no prior work in literature that investigates how individuals infer emotions from their own facial expressions, there is a great body of evidence documenting the ability of humans in accurately inferring emotion from the facial expressions of others [69, 173], even in cross-cultural contexts [192]. We thus expect that our participants will be able to infer their emotions when reviewing pictures of their own facial expressions.

(H2) Participants should be better able to infer their emotions when reviewing their pictures at the end of the day (Photo-Day Reconstruction), as opposed to one week later (Photo-Week Reconstruction). This hypothesis is based on the fact that emotional memories are reconstructed by drawing on information that resides in episodic memory [189]. However, at the end of the day there should be a significantly greater volume of contextual information one can draw on for reconstructing an experience and subsequently recall emotion, as opposed to one week later [11]. We thus expect participants at the end of the day to be significantly more accurate in inferring their emotions from a self-face picture that was captured a few hours before, than they would be from a picture that was captured a week before. Hence, we expect participants at the end of the day will be able to reconstruct and subsequently recall their emotions based on the contextual details captured in a self-face picture. A week later instead, we expect participants to infer their emotions from their self-face pictures simply by recognising their facial expressions.

(H3) Participants inferring their emotions based only on the time a self-face picture was taken (Time-Day Reconstruction) should in general be less accurate than participants inferring their emotions based on a self-face picture at the end of the day (Photo-Day Reconstruction). We expect this due to the added value of the self-face picture in the Photo-Day Reconstruction condition that can support both emotion recall and recognition. Nevertheless, the provision of timestamps in chronological order may assist the reconstruction process [124].

(H4) Relevant others should in principle be better able to infer emotion from the self-face pictures of our participants, given their increased exposure



(a) (b)

Figure 4.3. (a.) Clustering Areas of Interest (AOIs) for eye tracking analysis for each picture. (b.) Heat map produced by summarizing gaze behaviour on a self-face picture and the simple Lifelogging User Interface (LUI).

throughout daily life. This hypothesis is based on the fact that people are even more accurate in inferring emotion from facial expressions of familiar faces [130].

4.6.1 Measures

Motivated by previous work in the field of self-reporting on psychological well-being [44, 117, 132, 151, 197], we designed a simple Lifelogging User Interface (LUI) (Figure 4.1a) to inquire into participants' happiness at certain moments. By employing ESM, we asked participants to quantify their happiness using a continuous scale ranging from 0 (very bad) to 99 (very good). The same bar was also used during all the reconstruction sessions. For reducing the possibility that our participants would remember how they have previously used the bar for reporting their momentary emotions, we opted in for hiding the scale. The difference Δ between the self-reported emotion during experience sampling and during reconstruction signifies the participants' inaccuracy in reconstruction. A random sample of 10 rated pictures per condition (Photo-Day, Photo-Week, and Photo-Relevant) for each participant and his or her relevant other was chosen for eye tracking analysis. Each picture was preprocessed so that two major Areas of Interest (AOI) are defined: "Face" and "Background" (Figure 4.3). During eye tracking analysis, two metrics were measured:

- **Visit count** indicates the number of times a participant looked at a specified AOI.
- **Total visit duration** indicates the total time (seconds) a participant spent looking at a specified AOI.

In an attempt to understand how participants interacted with the LUI when asked to infer their emotion from their self-face pictures, four metrics were derived:

- **Photo duration** (in seconds) describes the overall time taken to evaluate a picture.
- **Photo duration** (in seconds) describes the overall time taken to evaluate a picture.
- **Events number** holds the total number of bar touches per picture.
- **Total events duration** (in milliseconds) describes the total duration of all bar-touch events observed.
- **Events number** holds the total number of bar touches per picture.
- **Total events duration** (in milliseconds) describes the total duration of all bar-touch events observed per picture.
- **Total delta** describes the total distance covered by the bar cursor during reconstruction.

Retrospective Think Aloud (RTA) [92] sessions were conducted to obtain qualitative insights into the way participants and their relevant others infer emotions from their self-face pictures and their relevant others' face pictures, respectively. RTAs were performed for all three conditions that include face pictures as **memory cues** (Photo-Day, Photo-Week, and Photo-Relevant). For this purpose, an RTA protocol was formed mainly questioning the rationale behind emotion inference.

4.6.2 Participants

Seven individuals (5 males, 2 females, median age 29 years) and seven relevant others, one for each individual (4 males, 3 females, median age 31 years), participated in the study for a total of two weeks. All were office workers with similar work patterns. They all used the application during working days. None of our participants suffered previous memory impairment.

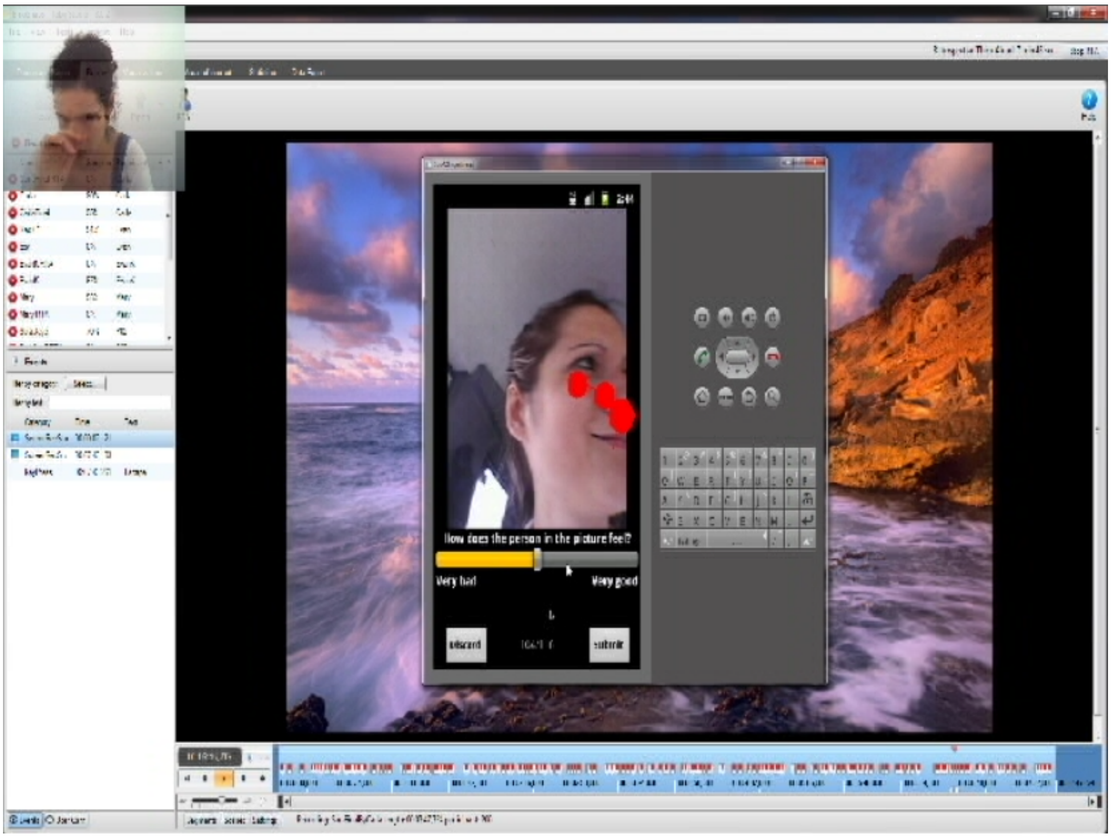


Figure 4.4. Interface used for performing Retrospective Think-Aloud (RTA) sessions.

4.6.3 Procedure

The study lasted two weeks in total. The first week was dedicated to Photo-Day (three days) and Time-Day (three days) reconstructions followed by a null day at the end. After a week had elapsed, we performed the Photo-Week and Photo-Relevant reconstructions. All participants were given a Nexus S mobile device with the EmoSnaps application pre-installed, and were asked to use it as their mobile phone. Each time a participant unlocked the screen, a self-face picture would be captured via the front–facing camera and the participant would be prompted to self-report on his or her psychological well-being using a validated single-item continuous scale [44] (Figure 4.1a). This would run for a total of six days (three days in Photo-Day and three in Time-Day, order counterbalanced across participants). One week after both Photo-Day and Time-Day reconstructions were completed, participants would perform the Photo-Week reconstruction, and their relevant others would perform the Photo-Relevant reconstruction. Pictures presented during both Photo-Week and Photo-Relevant reconstructions were captured during Time-Day reconstruction and hence, were not presented before. Each touch event on the bar indicating emotion was monitored, along with the time taken for each participant to evaluate each picture. A total of five out of seven participants repeated one Photo-Day reconstruction session and the Photo-Week reconstruction session on Tobii TX300 Eye Tracker, running an Android OS emulator at the size of Nexus S mobile device (Figure 4.4). Accordingly, a total of five corresponding relevant others also repeated the Photo-Relevant reconstruction session on the Eye Tracker with the same configurations. During the evaluation on the Eye Tracker, two synchronized video segments were captured, one screen video and one facial video. The screen video held all the actions a participant performed during the eye tracking session, while the facial video captured the participant's facial expressions. Upon ending, both participants and their relevant others went through a Retrospective Think Aloud (RTA) [92] session using as memory cues the two captured videos combined and presented in one interface (Figure 4.4).

4.7 Results

In this section, we present the results categorized according to the study research questions and hypotheses. The previously described measures are combined in an attempt to explain the observed phenomena.

4.7.1 Emotion Inference from Self-Face Pictures

A total of 584 pictures were captured in the course of the study. Participants and relevant others were able to infer emotions for approximately 70.6 % of the pictures (a). This finding is in-line with the our initial hypothesis (H1) in that participants would be in general able to infer their emotions from their self-face pictures. For the remaining 29.4 %, they selected the "Discard" option. As participants reported, this happened primarily due to poor lighting conditions, privacy concerns, incorrect posture, or inability to infer one's emotions from one's facial expressions.

"[P1] *I discarded it because it was blurry and poor. I wouldn't do it if the photo was looking silly, but I would do it for privacy reasons...*"
"[P2] *I am not really expressive in these pictures.*"
"[P4] *It's always the same! Looks like I don't have a happy face! It's a family problem I guess!*"

Discard rates ranged per condition with the highest discard rate observed a week after pictures were taken (Photo-Week, 36 %), followed by pictures reviewed at the end of the day (Photo-Day, 35.4 %), pictures reviewed by the relevant others (Photo-Relevant, 34.7 %), and timestamps of captured pictures reviewed at the end of the day (Time-Day, 2.3 %). A Pearson's chi-square analysis between the pictures reviewed a week after pictures were captured (Photo-Week), and the picture timestamps reviewed at the end of the day (Time-Day), on discard rates revealed a significant difference between the two distributions ($\chi^2(1, 659) = 99.83, p < .001$). In total, we obtained a sample of 1,002 valid pairs of emotion ratings, each pair consisting of an in-situ emotion evaluation coming from Experience Sampling and an emotion assessment coming from one of the four types of reconstruction sessions (Photo-Week, Photo-Day, Photo-Relevant, and Time-Day). On average, participants would capture a total of 15 pictures in a given day (min = 8, max = 29).

4.7.2 Context Recall vs. Facial Expression Recognition

An analysis of variance with the z-transformed[1] computed distance Δ between Experience Sampling and reconstruction values as dependent variable, and type of reconstruction (Photo-Day, Time-Day, Photo-Week, Photo-Relevant) as independent variable, displayed a significant main effect for the type of reconstruction ($F(3, 998) = 4.553, p < .01, \eta_p^2 = 0.014$). Post hoc tests using the Bonferroni correction revealed that participants assessing emotion a week after a self-face picture was captured (Photo-Week, $M = 9.722$, SD = 9.629) were significantly more consistent in estimating their emotion, as compared to assessing at the end of the day (Photo-Day, $M = 12.242$, SD = 11.857, $p < .05$) (Figure 4.5). However, no other significant effects were found. Nevertheless, this finding contradicts our hypothesis (H2) in that participants will be more accurate in inferring their emotions from their self-face pictures at the end of the day, as compared to a week later. Interestingly, participants drawing only on the timestamp a self-face picture was captured as a memory cue, were surprisingly accurate in inferring

[1] Z-transformation was applied to normalize the distance Δ between ESM and reconstruction ratings.

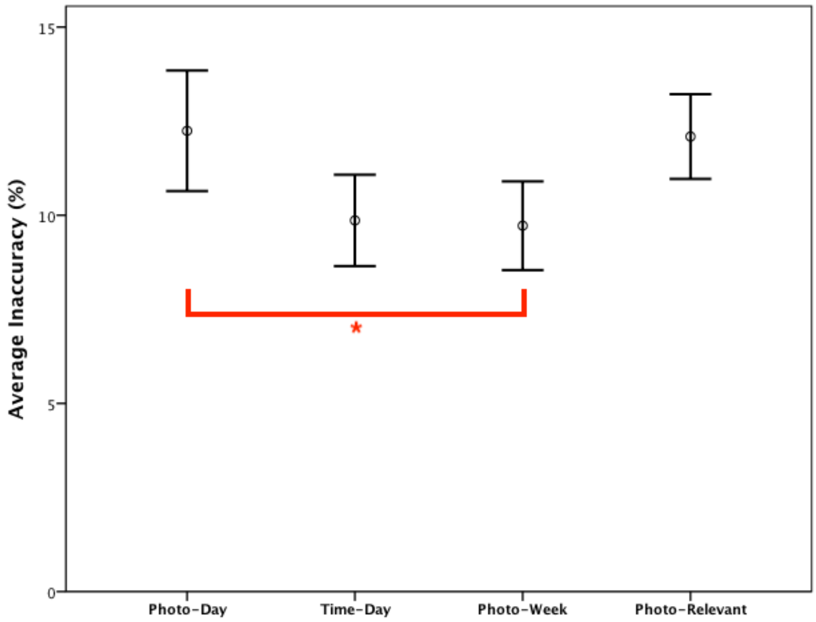


Figure 4.5. Z-transformed average inaccuracy per condition. Photo-Week was significantly more accurate compared to Photo-Day and Photo-Relevant.

their emotion. This is rather in conflict with our initial hypothesis (H3), where we assumed that participants during Time-Day Reconstruction will not be as accurate in inferring their emotions when compared to conditions that involved self-face picture review. Finally, these results also contradict our last hypothesis H4 in that relevant others would be more accurate in inferring emotion from our participants' facial expressions, than our participants themselves would be.

An analysis of variance with visit count (times eye gaze visits an area) and total visit duration (total time spent gazing at an area) as dependent variables, and type of reconstruction (Photo-Day, Photo-Week, Photo-Relevant) and AOI (Areas of Interest: Face and Background) as independent variables, displayed a significant main effect for the type of reconstruction and for the AOI on visit count ($F(2,72) = 4.251, p < .05, \eta_p^2 = 0.106$). Post hoc tests using the Bonferroni correction revealed that participants assessing emotions based on self-face pictures at the end of the day (Photo-day, $M = 1.240$, SD $= 2.067$) had significant higher visit count on the Background AOI than they had a week after a picture was captured (Photo-Week, $M = 0.320$, SD $= 0.627$, $p < .05$) (Figure 4.6), but no other significant effects were found. This indicates that at the end of the day, participants relied more on the context of the picture to infer their

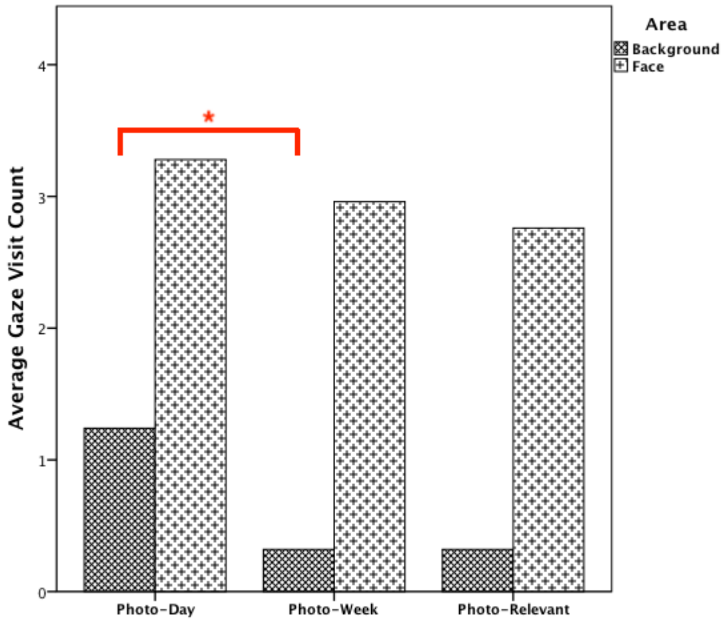


Figure 4.6. Average gaze visit count (average number of times eye gaze visits an area) per condition for Background and Face areas in a face picture. Face was proven the most dominant in eye gaze visits.

emotion than they did one week after a picture was captured (b). This finding is in-line with our initial hypothesis (H2) however, participants were significantly more accurate in inferring their facial expressions from a self-face picture a week later than they were at the end of the day.

Indeed, the majority of the participants assessing emotion based on self-face pictures at the end of the day repeatedly reported emotional inference primarily based on their location, co-presence, and/or the activity into which they were engaged, partially neglecting facial expressions (b):

> "[P2] *I know I was feeling pretty well because I was eating... You know that feeling when you are close to the tree and you eat more fruits than you actually eat at home.*"
> "[P5] *In this I know I was having lunch, because I know this is next to the bar. So I know I was feeling good because we were with Leonardo talking and making jokes so I know I was OK.*"

Such contextual information was derived from the background of each picture, when available, and was used as a memory cue to infer emotional state:

"[P4] *I can tell that because it is always a good time having breakfast with my colleagues all together, though it doesn't look like. I can say that it was breakfast time by the background.*"

One participant explained how he used contextual information to infer emotion through his self-face pictures:

"[P3] *I don't relate the context to emotion directly. I look at the context to recall what I was doing and by what I was doing I can recall if I was happy.*"

However, when the context remained the same during the day, it was reported of secondary importance:

"[P5] *I am pretty sure I took all the pictures at home so maybe the background is kind of secondary to me so I know where I was all the time.*"

Participants also reported that the demonstration of pictures in temporal order supported the process of inferring their emotion because it grants additional activity cues:

"[P2] *The sequence of the photos helps because I can understand what I was doing.*"

In contrast to inferring emotion from self-face pictures at the end of the day (Photo-Day), inferring emotion from self-face pictures a week after they were captured (Photo-Week) revealed an opposite effect. All participants reported emotional inference based on their facial expressions captured in self-face pictures. Facial expressions were preferred over context in multiple cases:

"[P3] *This one I cannot tell, am still at work from the context, but I don't see the mouth and I cannot really tell by the eyes so I discarded.*"
"[P4] *Here am smiling so I was feeling good, I only concentrate on my face that shows if you are happy or not and maybe the time but the face comes first.*"
"[P2] *I think the facial expression is essential for you to know if you feel OK or not, because if you go into the context it is always the same, hard to distinguish.*"

Participants also described the areas of their face on which they concentrated most during the reconstruction. Mouth, eyes, and eyebrows were the most referenced ones.

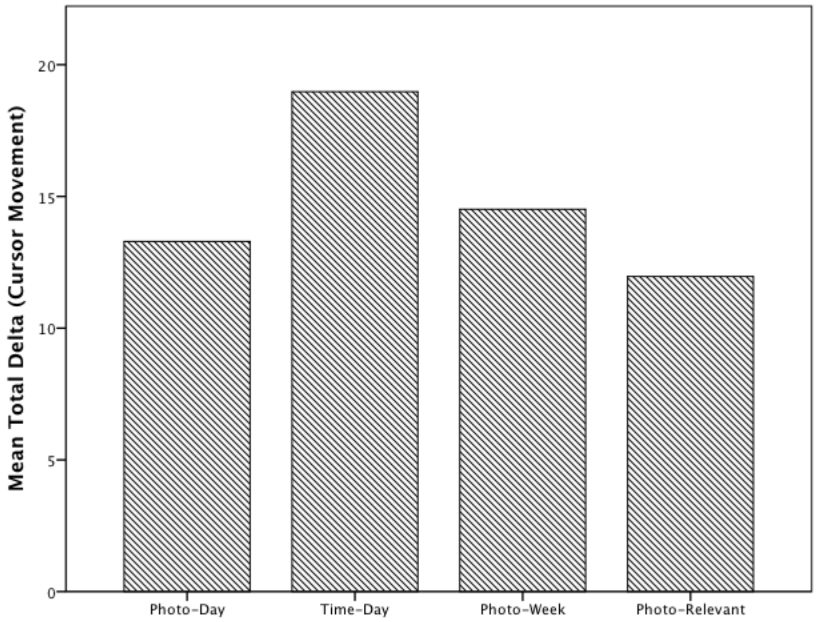


Figure 4.7. Average distance covered by cursor per condition (0–99) during evaluation. Time-Day entailed the highest bar cursor movement.

4.7.3 Relevant Other vs. Self

Post hoc tests using the Bonferroni correction revealed that the relevant others (Photo-Relevant, $M = 12.091$, SD = 9.59) were significantly less consistent, as compared to participants, in assessing participants' emotions from face pictures a week after they were captured (Photo-Week, $M = 9.722$, SD = 9.629, $p < .05$) (Figure 4.5) (c). Interestingly however, participants reviewing self-face pictures at the end of the day displayed a significant greater total delta than the relevant others did (Photo-Relevant, $M = 11.333$, SD = 6.813, $p < .05$) (Figure 4.7). This indicates that relevant others were significantly more certain when evaluating an individual's face picture than the individuals themselves were at the end of the day. This could be explained by the fact that relevant others ignored the context of the pictures they were evaluating. In fact, participants reviewing self-face pictures at the end of the day (Photo-Day) displayed a significant higher visit count on the Background AOI than the relevant others (Photo-Relevant, $M = 0.320$, SD = 0.556, $p < .05$) (Figure 4.6). This effect is complemented by the relevant others reporting that they relied totally on individual's facial expressions to infer emotion from pictures of familiar faces:

> "[R4] *I can't understand what she was doing by the pictures in none of them.*"

Inferring emotion in the absence of context was reported cumbersome in some cases:

> "[R3] *I have no indication I have no memory, he didn't come up in front of me to tell me how he was feeling so I try to guess, that way it makes it a lot harder.*"

Special facial expressions or grimaces were used as indicative of emotional state:

> "[R3] *So basically he is with his sunglasses on. He is pouting with his lips, this is something he does when he is not in a very good mood.*"

Similar to Photo-Week condition, specific areas on the face were used as cues to infer emotion. Mouth, eyes, and eyebrows are the most prevalent:

> "[R5] *So I look at 3 spots on the face: mouth, nose and between the eyebrows. More or less, when am not able to see all of them I discard.*"

4.8 Discussion

Overall, the results support our a priori expectation, in that participants would be able to infer their emotions when looking at their self-face pictures captured during mobile usage (H1). Surprisingly however, participants could more accurately infer their emotions when reviewing their self-face pictures one week after they were captured than they could at the end of the day (Figure 4.5). This contradicts our initial hypothesis (H2), in that participants would be more accurate when inferring emotion at the end of the day due to rich episodic memory recall. Eye tracking analysis revealed that when participants reviewed a self-face picture at the end of the day, they tended to gaze at the background more frequently than they did a week after a picture was captured. A possible explanation is that at the end of the day, participants repeatedly attempted to recall contextual information from the background in order to assess emotion in combination with their facial expressions. Moreover, participants reported that at the end of the day, contextual information derived from the background was used to infer activity and subsequently emotion. It is therefore probable that the process of recalling contextual information to assess emotion conflicted with the process of recognizing emotion through facial expressions and, thus, leading to reduced accuracy of emotional recall.

Surprisingly, participants reviewing self-face pictures a week after they were captured proved to be significantly more consistent in emotion assessment than their relevant others were. In both cases, a reluctance of context utilization can be assumed for different reasons. On the one hand, participants already experienced a one-week-long interval between capture and reconstruction, thus neglecting the context. On the other hand, context is meaningless for the relevant others as showed in the RTA results. Although both participants (Photo-Week) and their relevant others (Photo-Relevant) used the same areas of the face (mouth, eyes, and eyebrows) to infer emotion a week after a picture was captured, the relevant others proved to be less accurate. This contradicts our initial assumption (H4) that one's relevant others are better able to infer one's emotions from one's facial expressions, given increased exposure to them. One possible explanation for this is that even though z-transformation was applied to normalize the Δ between ESM and reconstruction ratings, relevant others did not have a notion on the scale used by the participants:

> "[R3] *I have no clue, he might have graded really happy or really sad and am not sure which grades he used.*"

Interestingly, the log data analysis revealed a significant higher bar cursor movement for participants evaluating their self-face pictures at the end of the day than their relevant others. This reveals a higher degree of uncertainty for participants inferring emotion at the end of the day than their relevant others. One reason is that relevant others simply knew that they were guessing, whereas participants tried to be as accurate as possible, incorporating any bit of contextual and/or facial information. Again, this finding supports the aforementioned approach of conflict between contextual detail recall and facial emotion recognition at the end of the day. Moreover, the considerable confidence of the relevant others attributes to the theory that humans accurately judge emotions from facial expressions, especially what happiness is concerned, at a precision level that yet cannot be achieved by algorithmic techniques [176]. Strangely, when participants assessed their emotions at end of the day (Time-Day) based only on temporal context (time a self-face picture was captured), these were found unexpectedly accurate (Figure 4.5), though not significantly more accurate. Also, the fact that in the same condition (Time-Day), participants exhibited significant greater total bar cursor movement in combination with the lowest discard rate (2.3 %) observed across all conditions lends credence to Day Reconstruction Method (DRM) [124]. More specifically, the increased bar cursor movement possibly indicates a background process of episodic recall based on subsequent temporal cues. This verifies that the disposition of subsequent events in temporal order can greatly support the recall of episodic memories [226], and therefore emotion. Unfortunately, no qualitative data are available for this condition, since it was initially designed as a control condition. Additional findings concern erroneous ways according which participants judged self-face pictures in order to decide about their emotional state. For example, female participants reported that the aesthetics of their self-face pictures influenced the way they were inferring their emotion, and in some cases, they had to discard the picture if they did not like it:

> "[P4] *I discarded it because it's awful!*"
> "[P4] *This one looks nice! The photo looks nice so I was feeling happy!*"
> "[P5] *That's the thing of being a girl again, I look at the picture and I am like oh I have such a huge nose! So am not sure it's kind of a girl thing but it is inevitable for me to not look at these kinds of things, sorry!*"

In our striving for capturing naturalistic behaviours, while promoting meaningful capture, we decided to employ transparent face picture capture during "screen unlock" on mobile devices. This provided a strategic opportunity to capture facial expressions under versatile mobile conditions. However, "screen unlock" is considered a procedural and rather mundane action performed when one wants to access one's smartphone. Thus, participants often reported lack of expressiveness in their self-face pictures mainly attributing to their own innateness:

> "[P2] *I am not really expressive in these pictures.*"
> "[P4] *However am not really expressive when am happy unless if am smiling, I have a serious face.*"

Moreover, some reported that the position of the mobile device during capture might have affected his ability to infer emotion from a face picture:

> "[R5] *The angle the picture was taken affects the image and adds shadow to several areas on the face like the eyes.*"

Interestingly, we had no privacy concerns reported, primarily because the study involved participants themselves and their close relevant others, or limited mobile device usage:

> "[P3] *Privacy concern? My data is not that much. I don't use the phone that much.*"

However, when participants were asked about reasons to discard a self-face picture, they mentioned privacy as a hypothetical reason:

> "[P1] *It was blurry and poor. I wouldn't do it if the photo was looking silly, but I would do it for privacy reasons.*"

In addition, visual cues held a surprisingly high amount of memory cues that relate to each other, maximizing the information they contain and thus triggering recall in explicit and implicit ways:

> "[P3] *Here I can see I was wearing my training jacket and I am probably going to the gym so I was feeling good.*"

4.9 Summary

This study inquired into *if* and *how* self-face pictures captured in mobile context could support users in inferring their day-to-day experiences, and more specifically the experience of happiness. We found that participants could better infer their emotions from self-face pictures one week following their capture, than at the end of the day. This was puzzling, as it contrasts established findings of episodic memory and common wisdom suggesting that memories dissipate over time. Our dominant hypothesis is that at any given emotional inference using self-face pictures, individuals could rely either on recall (i.e., a true recall of their episodic emotions that entailed, a. reconstructing of details from episodic memory, b. inferring their emotions, or more specifically happiness, from these episodic details), or recognition (i.e., direct interpretations of their emotions from their facial expressions, without much consideration of the context and the root cause of these). Some hours following capture, both these sources of information should be available – we expect that recognition is a more reliable route to emotion inference, yet individuals are likely to attempt to recall details from episodic memory and infer their emotions from these, thus introducing memory biases (e.g., confusing different locations or activities performed during their day). One week following capture, individuals are expected to have less capacity to recall episodic memories; thus, they rely more on the recognition route.

Generally, these findings provide support for the recognition approach, rather than the reconstruction approach. One possible explanation for this phenomenon could be that the process of inferring emotions from reconstructed episodic memories conflicts with the one of inferring them from facial expressions, thus disrupting the recall or recognition process. An alternative possible explanation could be a learning effect in the Photo-Week condition, as participants were more familiar with the reconstruction LUI, since they used it before in the Photo-Day condition. In any case, facial expressions, even when captured in mobile settings, were proven as good cues for inferring emotion, and thus they could be utilized for increasing the overall validity of retrospective self-reporting on (user) experience, as we will see in the next chapter. These findings corroborate the viability of our approach utilizing event-driven capture for producing memory cues, in the form of self-face pictures, that augment episodic memory recall (i.e., RQ1). Even as a proof of concept, the notion of event-driven capture produced memory cues with a tangible episodic memory benefit in the wild (i.e., cue-based augmented memory recall), and thus we took the next step and tested our approach in diverse everyday life settings and scenarios, as shown in the next chapter.

Chapter 5

Augmenting Memory Recall for UX Evaluation

In this chapter, we describe two studies that utilize our memory enhancing interventions for assessing quality of UX in the mobile and the automotive context. Our aim here is to demonstrate that event-driven capture can produce memory cues that substantially augment episodic memory recall, bearing significant practical potential in a wide range of daily–encountered contexts (i.e., RQ1). First, we present an extensive deployment of the previously showcased EmoSnaps prototype that attempts to assess its value in evaluating a wider set of mobile interactions and their associated emotions. Next, we briefly present a field study that assessed driver's UX by collecting self-reported anger and frustration levels before, during and after commuting. We then present a mobile application prototype[1] that utilizes cue-based augmented memory recall via event-driven memory cue capture, utilizing memory cues collected during a commute, for safely assessing driver's UX in a post-commute fashion. Finally, we discuss the potential of our interventions in substituting established self-reporting methods and we conclude.

5.1 Author's Contribution

The author of this thesis had a leading role in the Mobile UX study (Section 5.3) reported in this chapter. In particular, his contribution includes the conceptualization, development and deployment of the memory intervention, the study

[1] Author's contribution includes the conceptualization, design and development of the prototype. For more information, see the related publication [166].

design, and the data analyses, conducted in the frames of the original publication [165]. The co-author of the original publication provided his helpful guidance and knowledge on data analysis, while editing sections in the original publication. For more information, see the original publication [165].

In the Automotive UX study (Section 5.4), the author of this thesis had a secondary role, contributing partially in the conceptualization, fully in the development, and partially in the deployment of the experimental intervention. The co-authors of the original publication [251] lead the writing and analyses reported in the Automotive UX. Therefore, the Automotive UX study is briefly summarized, omitting elaborative statistical analyses and detailed description of results. For more information, see the original publication [251].

5.2 Introduction

Nowadays, the pervasiveness of modern (mobile) technologies renders User Experience (UX) increasingly important, since it drives user adoption and user satisfaction. There is a plethora of definitions about what UX is and why it matters. In their research agenda about User Experience, Hassenzahl and Tactinsky define UX as "*a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (complexity, purpose, usability, functionality, etc.), and the context within which the interaction occurs (organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)*" [105]. But how can one evaluate an innately highly–subjective and abstract notion such as UX? In fact, UX is not directly measurable thus, we resort to measuring UX aspects that are measurable as indicators of UX [101]. For example, efficiency is a performance-based indicator of UX that can be measured directly by eliciting a certain task that a user has to complete. In this case, a measure of efficiency could be task completion time. Additionally, questionnaires such as NASA-TLX [100], or the SUS [32] (see Chapter 3), provide good self-reporting indicators of other UX aspects (e.g., exhibited workload and perceived usability, respectively). UX as a phenomenon can be tracked and measured over time in a macro– and micro–temporal fashion [129]. For example, questionnaires administered at the end of a user task (e.g., SUS) typically aim at assessing UX aspects macro-temporally, while the Experience Sampling Method (ESM) is employed for assessing UX aspects micro-temporally (e.g., exhibited emotion right after one has sent an SMS). However, as we have previously seen, ESM (see Chapter 3) entails considerable drawbacks such as disrupting the user experience. In this chapter, we showcase the potential of our memory augmentation approaches to serve as effective alternatives to established methods of ecological momentary assessment (e.g., ESM) for assessing UX when established methods are highly disruptive (i.e., mobile context), or even dangerous (i.e., automotive context) to employ.

5.3 Mobile UX Evaluation

The study described in Chapter 4 provided promising results for the effectiveness of EmoSnaps, but it was limited to only one triggering event — the moment when users slide in to enable the screen of their device. With a second study, we wanted to inquire into the potential of EmoSnaps in capturing momentary emotions during a wider variety of interactions, such as when responding or initiating a phone call as well as, when launching different types of mobile applications. For this, we redeployed EmoSnaps in a week-long study with thirteen participants. We assumed that the increase in the range of events monitored would also result in capturing a wider range of associated emotions, as inferred from users' facial expressions. However, we expected some events to produce self-face pictures of higher quality than others, due to more appropriate posture and orientation of user's face against the front-facing camera of a mobile device. Thus, we wanted to establish which types of events EmoSnaps is most effective in monitoring.

5.3.1 Study Design and Procedure

In this deployment, EmoSnaps was installed on participants' own devices in order to increase the ecological validity of the study. Throughout the study, self-face pictures were being captured when one of the following events occurred: (1) screen unlock, (2) call answer, call end, or outgoing call, (3) SMS sent/read, (4) application launches, and (5) system actions. We found that the device played a significant role in the obtrusiveness of EmoSnaps. Some devices were able to capture a picture within 200 ms, while others required up to 800 ms, an effect that was noticeable to users and, at times, induced frustration. An upper threshold of 5 min was set at the capture frequency for each event, meaning that once a self-face picture was captured for a given event, no other self-face picture would be captured in the next 5 min relating to the same event.

Participants were instructed to review and rate their self-face pictures whenever they wished. We purposefully left the choice for when and how frequently to evaluate pictures to the participants in order to understand how they would behave in a real-life scenario: Would they perceive the task as a burden, or would they be intrinsically motivated to assess their emotions multiple times within the day?

5.3.2 Participants

Thirteen individuals (3 females, median age 28 years) participated in the study for one week. All were office workers with similar work patterns. They all used the application on their own Android devices.

5.3.3 Data Elicitation

Participants used a five-point Likert scale to respond to a single-item validated scale of psychological well-being "*How were you feeling at that time?*" [44, 132, 151] ranging from "*very bad*" to "*very good*". For each picture that the participants evaluated, we recorded the type of event that triggered the self-face picture capturing, the time of occurrence, and the time the picture was evaluated. At the end of the week, we conducted exit interviews to inquire into users' experiences with EmoSnaps.

5.3.4 Research Questions and Hypotheses

In order to validate our tool as a methodological instrument, we attempt to address the following four research questions:

(a) **Which interactions produce the greatest number of successfully emotionally assigned self-face pictures?** Our experience from the first deployment (see Chapter 4) showed that the quality of the captured pictures vary greatly depending on environmental conditions, such as illumination and user posture. Particularly, the increased diversity of mobile interactions monitored implies an arbitrary user posture in front of the device's camera, and thus, leading to an overall increase in the discard rate.

(b) **Can EmoSnaps reveal established patterns of the fluctuation of happiness over the course of a day?** Research in Psychology has shown that mood and perceived happiness fluctuate following diurnal and weekly patterns [194, 195, 217]. For example, morning hours are related to lower levels of happiness and higher levels of annoyance and anger [217]. Advocates of "Blue Monday" and "Weekend" effects claim that happiness levels are at a minimum on Mondays and increase during the week to reach a maximum on Fridays and weekends [194]. Accordingly, we expected that self-face pictures captured in the morning would be rated less happy than self-face pictures captured later in the day. Similarly, self-face pictures captured on a weekend would be rated happier than those captured during the week, especially on Mondays. Apparently, the perceived happiness elucidated from one's facial expressions is expected to still be influenced by external factors that are possibly unrelated to such temporal patterns (e.g., a bad day at work, or a pleasant conversation with a colleague).

(c) **Can EmoSnaps reveal meaningful differences on individuals' happiness over different activities?** Can this be captured via facial expressions and thus result in significantly happier rated self-face pictures? Given prior research [120], mobile communication occurs more frequently between individuals in relationships, and it is also used to increase family ties, maintain friendships, and provide mutual support. Similarly, social networking and instant messaging applications should follow the same norm [89]. Consequently, we expected that capturing self-face pictures in the context of mobile social interactions, such as calls or SMS, instant messaging, and social networking applications, would lead to pictures that would be rated significantly happier than pictures associated with other capture events.

(d) **Does happiness (or the lack of), as inferred from self-face pictures, correlate with an increase in mobile phone use?** Research has shown that mobile devices are habit-forming and can be a potential source of addiction, mainly because they provide quick access to rewards, such as social networking, communications, and news [174]. Accordingly, we expect some interactions to reveal patterns of use through triggering sequences of subsequent interactions. One would expect a generic interaction, such as screen unlock, to lead to a number of subsequent interactions; we assume that sessions that present a low number of subsequent interactions following a screen unlock have higher likelihood to represent habitual interactions. We assume such habitual interactions to be associated with decreased levels of happiness [174].

We also formed the following hypotheses:

(H1) Based on results presented in Chapter 4, we expect that EmoSnaps can capture a far wider range of events along with participants' exhibited facial expressions, than just "screen unlock" events. However, due to the diverse nature of these mobile interactions, we expect an increased rate of discarded self-face pictures.

(H2) We expect that emotion ratings as inferred from self-face pictures, as well as discard rates, will be systematically influenced by established and well-known happiness and mood fluctuations. For example, due to morning hours being related to lower levels of happiness and higher levels of annoyance and anger [217], we expect that self-face pictures captured in the morning will be systematically rated less happy than those captured later in the day. Similarly, self-face pictures captured in the weekend would be rated happier than those captured during the week, and especially on Mondays [194, 195].

(H3) We also expect that different types of mobile interactions will result in capturing facial expressions that reveal systematic differences in emotion. For example, we expect that self-face pictures captured during social networking and instant messaging applications will be rated significantly happier than others. We believe this is owed to mobile communication occurring more frequently between individuals in relationships, family members, and friends and thus, inducing a general happiness emotion [120] that could be captured by our intervention.

(H4) We assume that sessions that present a low number of subsequent interactions following a screen unlock have higher likelihood to represent habitual interactions. We assume such habitual interactions to be associated with decreased levels of happiness [174].

5.3.5 Results

Before presenting the results of the study, we define as events the mobile interactions that fall into the following five broad categories: (1) screen unlock, (2) call answer, call end (incoming/outgoing), and placing a call, (3) SMS sent/read, (4) application launches, and (5) system actions. In our system, each of these events would lead to the capturing of a self-face picture. In total, a set of 2,953 events and associated self-face pictures were captured from our 13 participants in the course of one week (approximately 32 events per day per person). Following previously proposed categorization approaches [27, 116], we group the events associated with self-face picture capturing in the following 22 categories sorted by frequency of occurrence:

- **Screen Unlock** (30.4 %): Although a systemic action, it was purposefully kept as a distinct category to relate to the findings of the previous study.
- **System** (18.2 %): Settings, Home, App Launcher, System UI, etc..
- **Calling** (16.1 %): Answering a call, placing a call, ending a call (incoming/outgoing), starting the dialer application, and initiating a contacts search.

- **SMS** (9.9 %): SMS/MMS read and sent.
- **Travel** (6.3 %): Google Maps, Maps, Waze, etc..
- **Social Networking** (4.5 %): Facebook, Twitter, LinkedIn, Instagram.
- **Web** (2.5 %): Android Browser, Chrome, Tunny Browser, Firefox.
- **Communication** (2.4 %): Skype, WhatsApp, gTalk, Viber.
- **Productivity** (2.2 %): Clock, Calendar, Memo, Notes, Menstrual Calendar, etc..
- **Other** (1.5 %): Unclassified apps.
- **Utilities** (1.5 %): Flashlight, Dictionary, Speedtest, Batterys, 3G Watchdog, etc..
- **File management** (.9 %): Astro, Mega, Dropbox, etc..
- **Image viewing** (.8 %): Gallery, Album, Infinite view, etc..
- **Entertainment** (.5 %): 9gag, Angry Birds, Simpsons, and other mobile games.
- **Google Play** (.5 %): Android vending.
- **Video Playing** (.4 %): YouTube, Android Video. Player, MX Tech video player.
- **Security** (.3 %): Avast, Clean Master, etc..
- **Weather** (.3 %): AccuWeather, Genie Widget, Weather Widget.
- **News** (.3 %): Pulse, Flipboard, etc..
- **Text Reading and Editing** (.3 %): Adobe Reader, Polaris viewer, Think Droid.
- **Music** (.2 %): Shazam, Jango mobile, etc..
- **New App Installed** (.1 %): Android Package Installer.

Discard Rates

All in all, participants were able to infer emotions for approximately 50 % (N = 1,477) of their self-face pictures (a); for the remaining 50 %, they clicked the "Discard" button. The observed increased discard rate compared to our first study (29.4 %) can possibly be explained by our attempt to increase the ecological validity in the second study, as initially hypothesized (H1). First, in this study as opposed to Chapter 4, a greater range of events was captured, some of which do not imply an appropriate posture. Second, the increased capture led to a higher number of rated pictures, which in turn might have tired the participants. Third, participants used their own mobile devices, which might have led to increased variance in the quality and timing of the captured pictures. For instance, we found that devices' speed in capturing a picture varied substantially (from 200 to 800 ms), which might have resulted in differences in captured posture.

Furthermore, different devices share different capabilities, producing in turn pictures of different quality in challenging conditions, such as ones of low-light or high-light exposure.

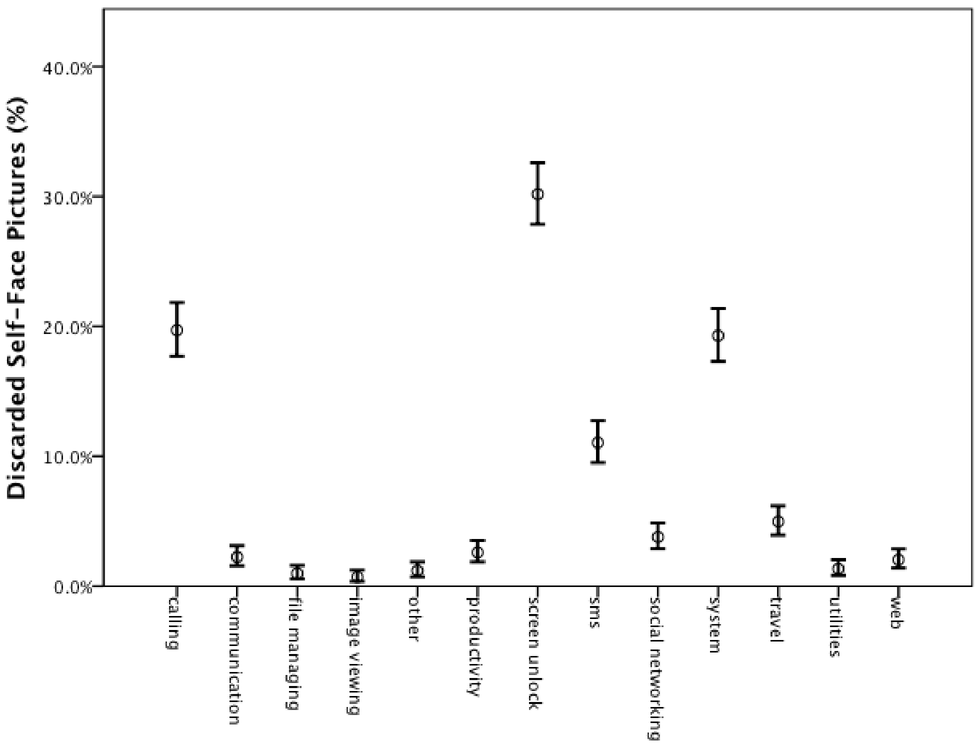
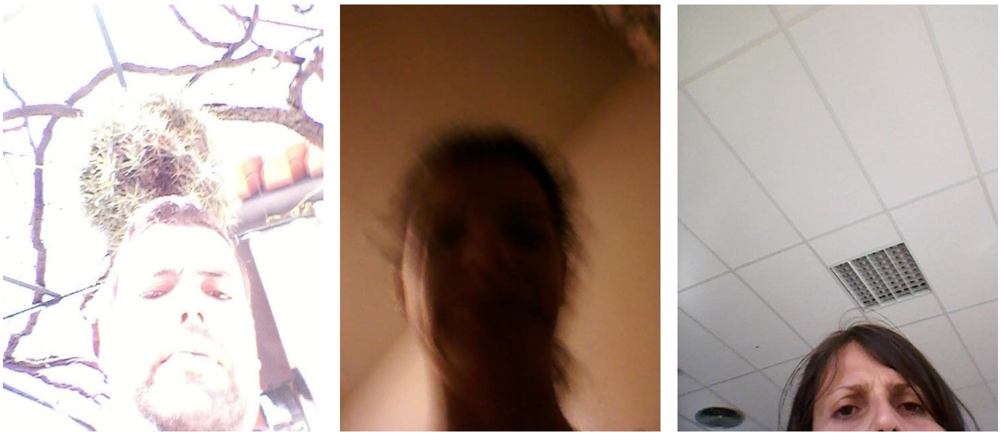


Figure 5.1. Ratio of discarded self-face generated for each event category. Categories with equal or lower than .5 % of total occurrence are excluded.

As expected, screen unlock was the most frequent sampled event (30.4 %), since it precedes any other interaction with a mobile device. However, 49 % of "screen unlock" events resulted in discarded self-face pictures, whereas in the previous study, the same event displayed a significantly lower discard rate (29.4 %) ($\chi^2(1, 1462) = 71.654, p < .001$) (Figure 5.1).

Events included in the "Calling" category led to the highest number of discarded self-face pictures (59 %), followed by "Productivity" (56 %), "SMS" (54 %), "File Managing" (54 %), and "System" (51 %). On the contrary, "Travel" and "Other" categories displayed the lowest number of discarded self-face pictures (38 % for each) followed by "Social Networking" (41 %), "Web" (42 %), "Image Viewing" (44 %), and "Communication" (46 %). The increased discard rate observed in some categories (e.g., Calling and SMS) can be attributed to the type



(a) (b) (c)

Figure 5.2. Some examples of self-face pictures that participants discarded. In (a.), the luminosity is high; in (b.), the camera focus time is insufficient; and in (c.), the participant's face is not fully included in the picture.

of interaction these categories imply. Incorrect posture of the face in front of the mobile device, or insufficient capturing time, may affect the overall quality of self-face pictures being captured (Figure 5.2). For example, when answering a call, the participant quickly grabs her device, presses the "answer" button, and holds it next to her ear, whereas when browsing the Web, the device is held in a stable position in front of the face, resulting in self-face pictures of greater quality.

In order to understand better this phenomenon, we took a closer look inside "Calling" and "SMS" categories, and particularly their sub-events. The "call answer" event, included in "Calling" category, systematically produced the highest number of discarded self-face pictures (92 %), in contrast to "end incoming call" event (48 %, $\chi^2(1,105) = 22.537, p < .001$), "call answer", and "end outgoing call" ($\chi^2(1,115) = 35.734, p < .001$). The same effect is again observed among events included in the "SMS" category with "SMS read" producing a 44 % of discarded self-face pictures, whereas "SMS sent" displayed 67 % ($\chi^2(1,97) = 4.222, p < .05$). Again, this effect can be attributed to insufficient exposure time of the face in front of the camera after the "SMS sent" event occurred, leading to lower quality of self-face pictures (Figure 5.2).

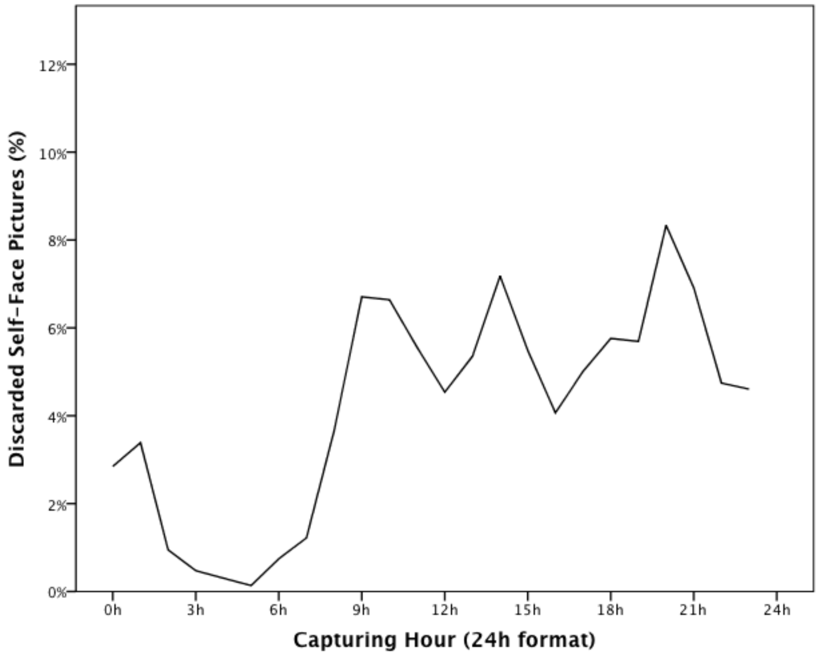
As expected, time of the day had an impact on discard rates, with self-face pictures captured during daytime displaying lower discard rate than those captured during night-time (H2). More specifically, self-face pictures captured at 17:00 (39 %) and at 11:00 (42 %) displayed the lowest discard rates, while those captured at 23:00 (69 %) and at 21:00 (71 %) displayed the highest (Figure 5.3a). During the interviews and by visually inspecting these pictures, we confirmed that these time effects can primarily be attributed to luminosity variation between daytime and night-time, which has a strong influence on the quality of the captured self-face pictures (Figure 5.2).

We also found that the time at which individuals reviewed their pictures had an impact on discard rates. Self-face pictures evaluated early in the morning and in the afternoon showed a lower discard rate, varying between 35 % for pictures evaluated at 07:00 and 38 % for pictures evaluated at 17:00. In contrast, pictures reviewed at the end of the day had the highest discard rate, up to 80 % at 23:00 (Figure 5.3b). A Pearson's chi-square analysis between discarded and rated distributions for the above time frames revealed a significant main effect between 07:00 and 23:00 ($\chi^2(1,112) = 21.063, p < .001$) and between 17:00 and 23:00 ($\chi^2(1,226) = 37.264, p < .001$). We attributed the above phenomenon to the effect of tiredness that accumulates during the day and reaches its maximum late at night [217], influencing the participants' ability and will to review their self-face pictures. A quick glimpse on the number of self-face pictures evaluated daily reveals a maximum on Tuesday (62 %) and on Monday (53 %), in contrast to the lowest rates exhibited on Sunday (36 %) and Saturday (37 %) (b). A Pearson's chi-square analysis between discarded and rated distributions for the aforementioned days revealed a strong significant main effect between Tuesday and Sunday ($\chi^2(1,684) = 40.998, p < .001$), and between Monday and Saturday ($\chi^2(1,724) = 18.193, p < .001$). One explanation could simply be that mobile interactions are more frequent during working days than they are on the weekend [234].

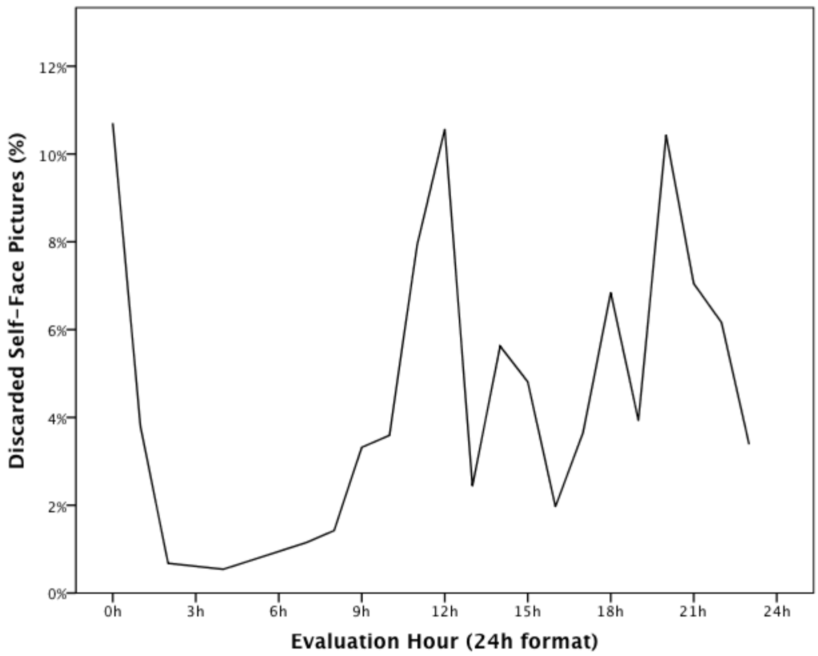
Happiness over Time

A glimpse at the hourly variation of happiness, as reported based on captured self-face pictures, reveals distinct patterns for mood fluctuations during both weekdays and on the weekend (b). While no pattern can be observed during weekends, during weekdays, happiness seems to display a low in the early hours of day, as previously hypothesized (H2), with a constant increase over the course of the day (Figure 5.4a).

An analysis of variance with happiness ratings as dependent variable and the hour of the day as independent variable for weekdays and the weekend displayed a significant main effect for the hour of the day, as far as weekdays are concerned ($F(201,285) = 3.741, p < .001, \eta_p^2 = .056$) (Figure 5.4a), but not for the weekend. Post hoc tests using the Bonferroni correction revealed that self-face pictures

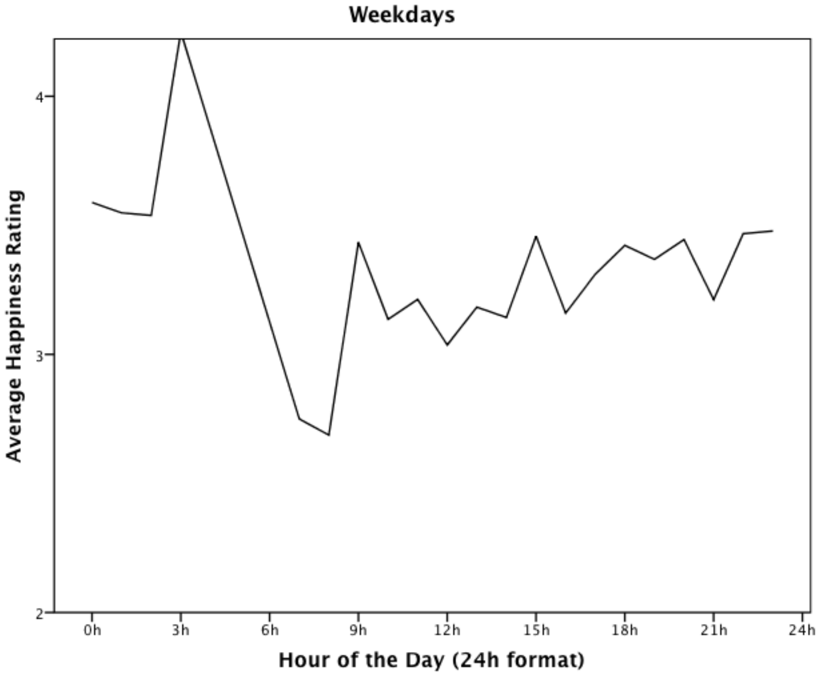


(a)

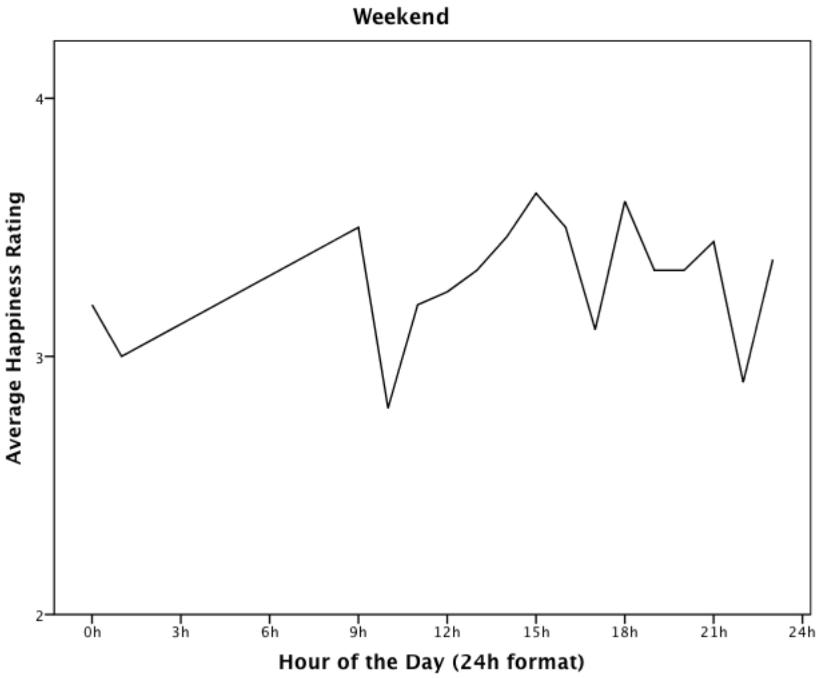


(b)

Figure 5.3. Ratio of discarded self-face pictures in relation to (a.) the hour they were captured and (b.) the hour they were evaluated.



(a)



(b)

Figure 5.4. Average happiness rating fluctuation observed during (a.) weekdays and (b.) weekends, as reported based on self-face pictures captured.

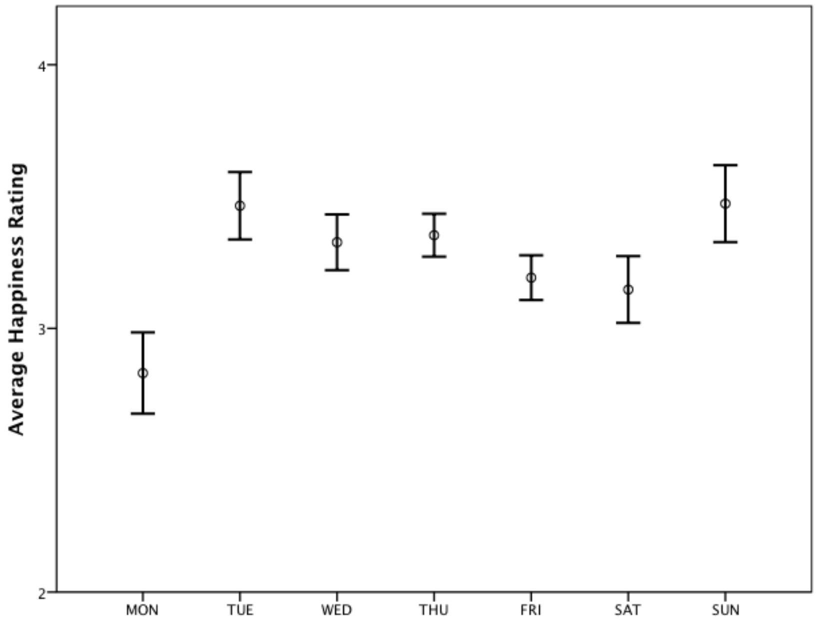


Figure 5.5. Average happiness rating per evaluation day for all self-face pictures.

particularly captured at 08:00 ($M = 2.688$, SD = .143, $p < .05$) were rated significantly less happy than pictures captured during almost the rest of the day. The other way around, self-face pictures captured late at night (01:00 to 03:00) and particularly at 03:00 ($M = 4.25$, SD = .286) were found to be significantly happier than pictures captured at 08:00, 10:00 ($M = 3.136, SD = .072, p < .05$), 12:00 ($M = 3.037, SD = .089, p < .05$), and 14:00 ($M = 3.143$, SD = .082, $p < .05$) (b). No other significant effects were found. These results are in line with existing findings in the psychology of well-being and our initial hypothesis (H2), suggesting that daily frustrations, such as early wake up, coordination of family activities, and daily commute, contribute to negative feelings, and that a constant increase in happiness is observed over the course of a weekday [217].

Systematic effects are also observed in the variation of happiness over the course of the week. An analysis of variance with the happiness rating as dependent variable and the day of the week a self-face picture was evaluated as independent variable displayed a significant main effect for the weekday ($F(61, 476) = 7.68, p < .05, \eta_p^2 = .03$) (Figure 5.5). Post hoc tests using the Bonferroni correction revealed that pictures evaluated on Tuesday ($M = 3.465$, SD = .887, $p < .05$), Wednesday ($M = 3.326$, SD = .839, $p < .05$), Thursday ($M = 3.353$, SD

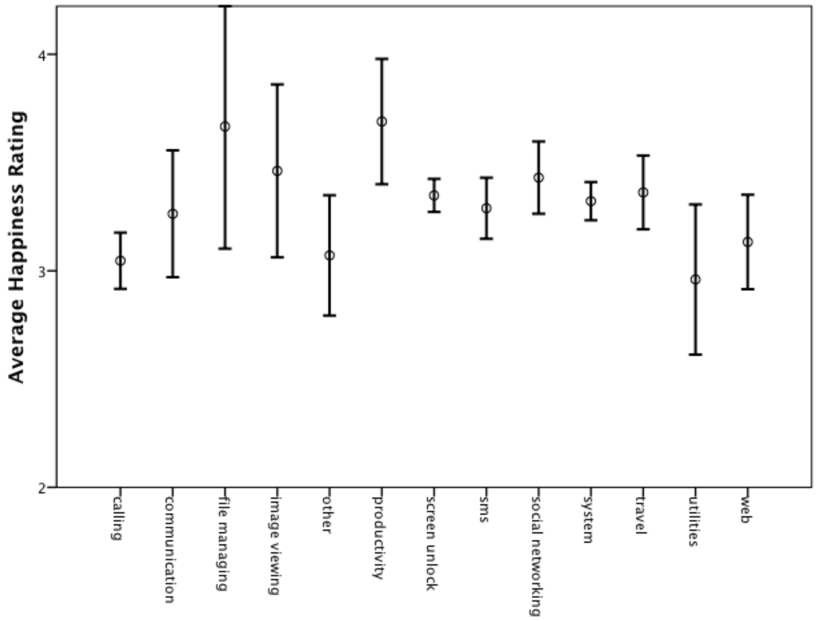


Figure 5.6. Average happiness per event category as reported based on self-face pictures. Categories with equal or lower than .5 % of total occurrence are excluded.

= .675, $p < .05$), Friday ($M = 3.192$, SD = .892, $p < .05$), and Sunday ($M = 3.473$, SD = .844, $p < .05$) were rated significantly happier than pictures evaluated on Monday ($M = 2.83$, SD = .591). In addition, pictures evaluated on Sunday were also rated significantly happier than pictures evaluated on Saturday ($M = 3.147$, SD = .818, $p < .05$) and Friday ($M = 3.192$, SD = .892, $p < .05$), but no additional significant effects were found. These results corroborate the well-known "Blue Monday" and "Weekend" psychological effects [194, 217]. Overall, EmoSnaps seems able to capture the daily and weekly variation of happiness (H2), as reflected in the current psychological literature.

Happiness across Interactions

Next, we looked at the effect that various mobile interactions have on one's happiness. In this analysis, we excluded all event categories that reflected less than .5 % of the total number of sampled events.

An analysis of variance with the happiness rating as dependent variable and the type of event category as independent variable revealed a significant main effect for the type of the event category ($F(12, 1476) = 2.278, p < .001, \eta_p^2 = .027$) (Figure 5.6). Given prior research, one would expect that events related to phone calls and writing or reading short messages (SMS) to be associated with higher levels of happiness, as they reflect individuals' social interactions, an inherently joyful activity [57, 234]. Surprisingly though, we found the exact opposite with respect to the category "Calling" (Figure 5.6). Post hoc tests using the Bonferroni correction revealed that self-face pictures captured through events falling into "Productivity" ($M = 3.689$, SD = .76 $p < .01$), screen unlock ($M = 3.348$, SD = .84, $p < .005$), "Social Networking" ($M = 3.43$, SD = .745, $p < .05$), and "System" ($M = 3.321$, SD = .725, $p < .05$) categories were rated significantly happier than self-face pictures captured when "Calling" ($M = 3.046$, SD = .914) events occurred (c). No significant differences were found for the "SMS" category. One possible explanation for this phenomenon could be the intrusion effect that a call implies, leading to disruptiveness of the current task or social interaction [57]. One would expect outgoing calls to not display the same effect, but no significant differences were found between self-face pictures captured during incoming and outgoing calls. Similarly, "Calling" events occurring during the weekend were expected to produce happier assigned self-face pictures, since weekends are associated with non-work activities and greater well-being [89, 120, 195]. However, no sufficient proof was found that "Calling" would result in happier emotionally assigned self-face pictures on the weekend.

Another interesting finding is that "System" and "Productivity" related events and applications were found to produce significantly happier self-face pictures than "Calling" did (c). This could potentially be explained by the feeling of control over one's device ("System") and one's life ("Productivity") evoked by these applications. Increased feeling of control is positively related to increased happiness [141], and thus we can assume that self-face pictures captured during the use of applications that provide control and support scheduling were systematically rated as happier than those captured when calling. On the other hand, the "Social Networking" category was found to support our initial hypothesis (H3) that social networking applications will produce happier self-face pictures.

Interestingly, the "screen unlock" event displayed very small standard deviation (SD = .84, N = 465) compared to other event categories (Figure 5.6), and we also found it to produce significantly happier self-face pictures than the "Calling" category did. This indicates that during a screen unlock, participants displayed similar facial expressions that they systematically rated happier than average ($M = 3.283$, SD = .835). Indeed, a one-sample t-test revealed a significant difference between the overall mean reported happiness (excluding "screen unlock" ratings) and mean reported happiness of "screen unlock" events between the two distributions ($t(464) = 2.422, M = .094$, SD = .084, $p < .05$). From a user experience perspective, the screen unlock is considered a neutral interaction. However, it is a rather important event as it signals the beginning of further interaction with a mobile device. Therefore, a self-face picture captured when a screen unlock event occurs is expected to display expressions that are mainly induced by the current affective state of the user and yet remain irrespective of mobile device use. This potentially indicates that EmoSnaps could capture the nuances of everyday life.

Habits Emergence

Recent research has shown that a standard mobile use session lasts approximately less than a minute [27, 174]. In pursuit of revealing additional behavioural insights, we clustered all occurring events in (overlapping) time frames with duration of 2 min each. For each trigger event and corresponding self-face picture, we examined the number of preceding and subsequent events that occurred within these 2 min, and compared these events with the reported happiness based on the corresponding self-face picture. By investigating preceding and subsequent events, we attempted to reveal plausible effects for the reported happiness, such as the impact of motivational orientation on happiness and how this carries over to subsequent interactions.

At first, we inquired into the potential interaction of happiness with frequency of use and particularly whether an increase in frequency of entailing events (both preceding and subsequent) would indicate also an increase in happiness, as reported via self-face pictures (d). This was intended to unveil the potential engagement to emotionally rewarding habits in case it could be displayed on self-face pictures. An analysis of variance with the happiness rating as dependent variable and the number of preceding events occurred for each event, within a 2-min period, as independent variable revealed a significant main effect for the number of preceding events ($F(8, 1430) = 3.001, p < .005, \eta_{\mathrm{p}}^2 = .017$). Accordingly, post hoc tests using the Bonferroni correction showed that self-face pictures of events with no prior event were rated significantly happier (M = 3.38, SD = .847) than pictures of events that followed after 3 consecutive events (M = 3.02, SD = .855, $p < .005$) within a 2-min period (Figure 5.7), but no further significant effects were detected. This result indicates that participants' reported happiness based on self-face pictures reaches greater levels when the number of interactions remains limited (H4). However, the same analysis for happiness rating and subsequent events revealed no significant main effect ($F(8, 1430) = .510, p < .05, \eta_{\mathrm{p}}^2 = .003$). At this point, the top five most occur-

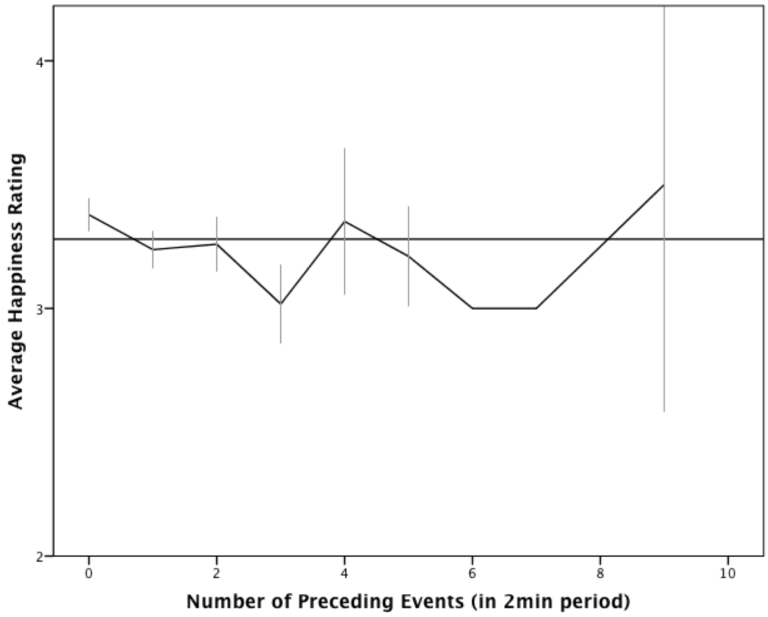


Figure 5.7. Average happiness rating per number of preceding events within 2-min period. Pictures of events with no preceding event were rated significantly happier than pictures of events with 3 preceding events.

ring event categories were selected and examined separately, with respect to the reported happiness versus the frequency of entailing events, but no significant main effects were detected.

Next, we investigated the frequency of preceding and subsequent events per category. A multivariate analysis of variance with the number of preceding and subsequent events occurred per event, within a 2-min period, as dependent variables and the type of event category as independent variable revealed a significant main effect for the type of the event category on the number of preceding ($F(12, 2857) = 92.986, p < .001, \eta_p^2 = .281$) and the number of subsequent ($F(122, 857) = 23.594, p < .001, \eta_p^2 = .09$) events. Not surprisingly, post hoc tests using the Bonferroni correction revealed that the screen unlock event systematically displayed the minimum number of preceding events ($M = .06$, SD = .239, $p < .001$) than all other event categories did. This is explained by the fact that screen unlock comprises the very first action that a user has to perform in order to start interacting with a mobile device. Thus, the probability of detecting events prior to screen unlock is low even in a 2-min period. In contrast, screen unlock systematically entailed the maximum number of subsequent events (M = 1.58, SD = 1.438, $p < .001$) within a 2-min period compared to all other

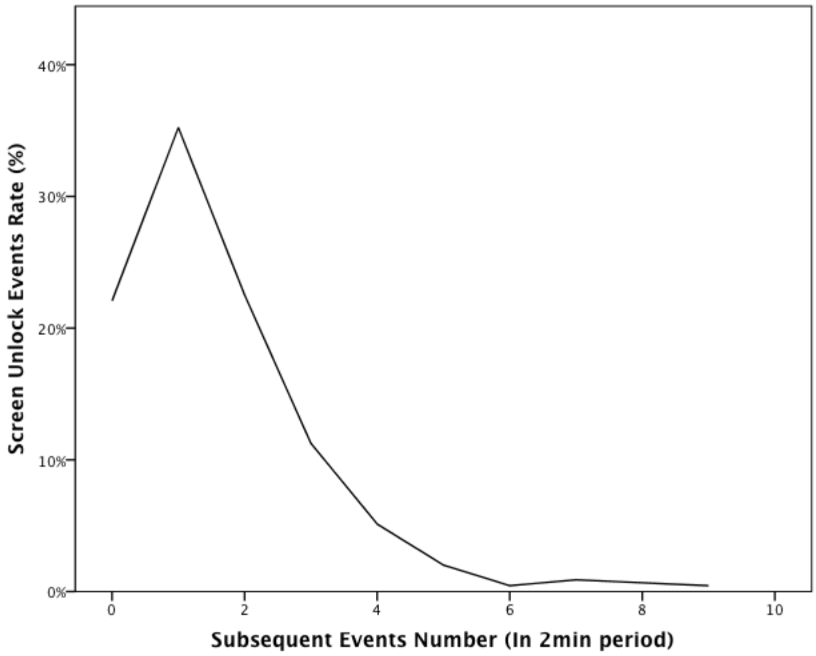


Figure 5.8. Screen unlock events rate in relation to number of events they entailed, within a period of 2 min.

event categories, apart from "File Managing" category ($M = .81$, SD = .849, $p < .05$). However, for the 22.1 % of the times that participants unlocked their mobile device, no event was detected for the next 2 min (Figure 5.8). For the rest 35.2 % and 22.5 % of screen unlock events, one and two subsequent events were recorded, respectively, within a period of 2 min (Figure 5.8). Overall, taking into account that the screen unlock event was found by far the most frequent action performed with a mobile device (30.4 %), occurring in average 1.76 times per hour ($min = 1$, $max = 6$, SD = .914), strong indications arise for a checking habit formation on behalf of the participants [174]. In other words, users were most of the times unlocking their mobile device to check something (time, missed calls, SMS, e-mails, etc.), but without engaging in lengthy interactions with it. This finding is supported by the fact that mobile use sessions typically last less than a minute [27, 174]. The built-in functionality of Android task bar can potentially explain the observed phenomenon or satisfy the checking habit need, since it supports fast and easy access to a wide range of notifications (e-mails, SMS, missed calls, Facebook updates, etc.).

5.3.6 Established Findings

Overall, this work corroborated findings from studies on psychological well-being, which demonstrates that EmoSnaps can be used in measuring users' emotions and experiences during everyday life in mobile context. For instance, we were able to detect diurnal and weekly variations in mood attributed to factors such as daily and weekly routine (H2). Despite the increase in discarded self-face pictures, as compared to the first study, the discard rate of 50 % reflects an acceptable level for a real-life study. Interestingly, we found participants to review their self-face pictures more frequently than we expected, while they systematically rated them above the expected average of 3 ($M = 3.283$, SD = .835). Most participants reviewed them on a daily basis, with some participants performing the task multiple times during the day. Participants often reported that the tool offered them personal value, as it enabled them to review how their emotions vary over the course of a day, and provided them with an activity to perform during idle periods of time.

As participants were all office workers with similar working patterns, we expected them to follow a similar diurnal routine. This was confirmed by the number of events occurring hourly during weekdays. The maximum number of mobile interactions was found in the morning during wake up and commuting. Interesting insights were revealed from the perspective of happiness variation throughout the day and the week. In agreement with our initial assumption, morning hours (08:00) displayed a daily happiness minimum, whereas self-face pictures captured late at night (03:00) revealed a happiness maximum. Similarly, as expected, self-face pictures captured on Monday were rated significantly less happy than almost all the self-face pictures captured during the rest of the week. These results are aligned with the psychology of well-being and the known impact of daily hassles on one's happiness levels [194, 217]

Next, we were surprised to discover that social interactions, and particularly calling, contribute negatively to individuals' happiness. This contradicts our a priori expectations, in that mobile social interactions, such as SMS and Calls, as yet another form of social interaction, should lead to increased happiness. However, we believe that the observed phenomenon can be attributed to the effect of intrusion that an incoming call may imply [57]. Yet, no sufficient evidence was found to justify why the phenomenon was observed in outgoing calls as well. However, "Social Networking" was found to support our initial hypothesis (H3) that it increases happiness, at least as inferred from individuals' facial expressions. In addition, "Productivity" and "System" events were rather surprisingly also associated with increased levels of happiness. One plausible justification

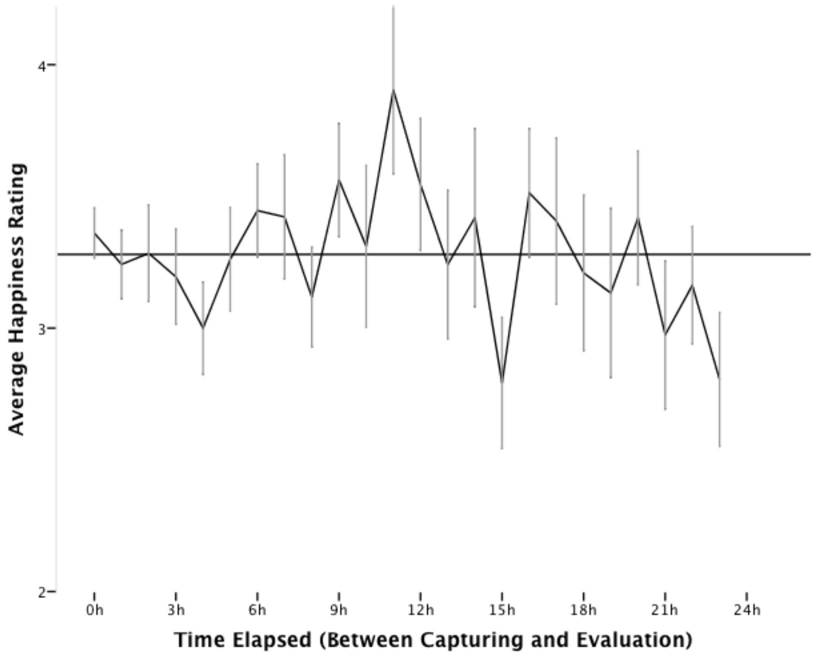


Figure 5.9. Average happiness rating in relation to time elapsed between capturing and evaluation.

might be the increased feeling of control that these types of mobile interactions induce on individuals [141].

Interestingly, our findings confirm prior insights into the habitual use of mobile devices [27, 174]. More specifically, we found participants to frequently "slide in" to access the functionality of their mobile device without engaging into further interactions. We perceived this as a habitual interaction when individuals access their mobile device to check current status, such as missed calls, e-mail, or other kind of notifications. This particular habit is thought to be rewarding and, thus, increases overall mobile use [174], when in fact participants were found to check their mobile devices approximately 1.76 times per hour. However, we were not able to prove our initial assumption, in that habitual interactions are associated with decreased levels of happiness (H4), since the number of subsequent interactions displayed no significant effect on happiness levels, as reported based on self-face pictures. Moreover, a tendency was observed on participants' ratings to be located very close to global reported happiness average ($M = 3.283$, SD $=$ .835), when less than 2 h had elapsed between capturing and evaluation (Figure 5.9). Subsequently, happiness ratings displayed a local minimum 4 h after the capturing, to start increasing again and reveal a global maximum when 11 h had elapsed since capturing. A global minimum was observed when 15 h had passed between capturing and evaluation. The displayed variation on reported happiness in relation to the temporal difference between capturing and evaluation might indicate an attempt to reconstruct an experience based perhaps on the context that a self-face picture holds, given the fact that incidents remain in episodic memory for such small time intervals [226].

Some mobile interactions were easier to capture than others, mainly due to posture of the face in front of the camera of the mobile device and the exposure time that each event involved. For example, "Web", "Travelling", and "Social Networking" events were easy to capture, whereas "Calling" and "SMS" proved somehow cumbersome, leading to higher discard rates. Environmental factors played a significant role in the quality of the self-face pictures captured. It was shown that pictures captured at night revealed a higher discard rate than pictures captured within the day. Moreover, the time a self-face picture was reviewed affected the outcome of the evaluation. Pictures evaluated late at night were more prone to discarding than pictures evaluated earlier in the day, potentially due to participants' increased tiredness at the end of the day [217]. Similar to the previous study, no participant raised any privacy concerns regarding the capturing of self-face pictures. As expected, knowing that self-face pictures are only locally stored was crucial to participants. Yet, some participants raised concerns over the use logs, as this process was less transparent to them.

The first EmoSnaps study, presented in Chapter 4, was very informative as it demonstrated that using EmoSnaps, individuals (1) were able to infer their moment-to-moment emotions remarkably reliably, and (2) even more interestingly, the most effective path to performing this task was through a combination of emotion reconstruction (via episodic recall) and facial expression recognition, whose ability does not decay with time. However, this study did not prove EmoSnaps' ability to measure users' feelings induced by specific applications but rather their usage-independent levels of happiness. The second study opted to understand how using different kinds of applications on our smartphones affect our happiness. We thus sampled users' facial expressions triggered by a wider set of system events, such as receiving phone calls and accessing different types of applications. The study revealed significant differences in users' happiness across different kinds of uses with the smartphones. Interestingly, social interactions such as receive a phone call were associated with reduced levels of happiness, while productivity applications were associated with increased levels of happiness. Moreover, we found systematic variations of happiness over the course of a day as well as the week, which were largely in agreement with the established findings in positive psychology. All in all, the results of both studies provided us with confidence over the effectiveness of event-driven capture in producing memory cues that augment episodic memory recall (i.e., RQ1).

5.4 Automotive UX Evaluation

Nowadays, with autonomous cars still in their infancy, driving is a well–known activity for its capacity to elicit primarily negative emotions and feelings (e.g., anger, frustration, boredom, etc.). Therefore, the automotive context poses an unparalleled opportunity for trialling the effectiveness of event-driven capture in augmenting episodic memory recall in practice (i.e., RQ1), for supporting the recall of momentary emotions while driving, but in a retrospective fashion. In fact, the driver's emotional state is an important issue for automotive safety [73]. A number of driving behaviours are negatively affected by emotions, linking anger or aggression to accidents (e.g., [139]). Prior work points out the importance of anticipating driver frustrations in order to increase road safety [97]. Taib et al., [220] attempted to study frustration detection with posture sensors in the car and helped to detect dangerous levels of frustration, focusing on the general experience of commuting. Frustration can be distinguished from anger by the degree of negativity and arousal. Being closely related to frustration, anger is also a regularly occurring phenomenon with regard to driving behaviour (e.g., [215]). Anger has been associated with a sense that the self has been offended or injured, with the belief that another person was responsible for the event [214]. In summary, anger is related to a specific source often linked to a social context (e.g., "*the driver behind is pressing me*"), whereas frustration is more unspecific than anger, often linked to environmental circumstances that impede a person's actions (e.g., red traffic lights).

For investigating how people predict, experience and recall anger and frustration before, during, and after commuting, we developed the eMotion application. The eMotion application enables the capture of actual experience throughout commuting, employing the Experience Sampling Method (ESM), an in-situ method that allows the collection of data as an experience unfolds without a researcher being present. In particular, we applied event-driven capture, as our focus was on particular discrete events in participants' lives, e.g., traffic congestion. The data collection was done via smartphones with the eMotion application installed. The eMotion application prompts drivers for emotional self-reporting at specified moments (e.g., when the car has just stopped or when it is moving slower than 5 km/h). We used the eMotion application not only to capture the data during a commute, but as an overall data collection tool for anger/frustration prediction and recall before and after commute. Additional data was captured, such as average speed, distance covered, time, and duration of a commute. The eMotion application was available in two versions, German and English. All collected data was anonymized and automatically uploaded from the eMotion application to an online MySQL database whenever an Internet connection was available.

5.4.1 Field Study

We deployed the eMotion application in a field study particularly aiming for answering the following research questions:

i. How does the level of exhibited anger and frustration change with regard to different phases when commuting (i.e., predicting, experiencing and recalling)?
ii. How additional factors such as mood, time of the day, and predictability of congestions relate to commuters experienced levels of anger and frustration?

For answering our research questions, we formed 3 conditions:

A. **Predict.** In this condition, participants were asked to predict their anger and frustration levels by the time they would have reached their destination before they start commuting. Participants were also asked to report their anger and frustration levels also when stopped at traffic during a commute.
B. **Experience.** In this condition, participants were only asked to report their anger and frustration levels when and if stopped at traffic.
C. **Recall.** In this condition, participants were asked to recall their anger and frustration levels after they had reached their destination in case there was a congestion. Participants were also asked to report their anger and frustration levels also when stopped at traffic during a commute.

We also formed the following hypotheses:

(H1) We expect that when participants recall a prior negative commuting experience they will report less frustration and anger levels, as opposed to when experiencing or predicting it, due to the manifestation of the "rosy view" effect [161]: negative past experiences are considered as less negative when sufficient time has elapsed between memory encoding and recall.
(H2) We expect additional factors, such as mood and unexpected congestions, will have a significant impact on commuters exhibited anger and frustration levels.

Figure 5.10. Mounted mobile device running the eMotion experience sampling mobile application.

Overall, we collected data from 10 commuters (4 in Portugal and 6 in Austria) over the course of three weeks, with each week covering one condition in the daily commute: A. Predict, B. Experience, and C. Recall. Every participant received a mobile phone with the eMotion application pre-installed, along with a wind-shield mobile device holder, and was instructed how to set and use the device. Upon departure, participants were instructed to select one of the three available options depending on their individual study plan. In option A (predicting), participants were asked to estimate how angry and frustrated they will be at the end of their commute. Contrary to A, option C (recalling) asked participants upon arrival to their destination to recall how angry and frustrated they were during the trip. In option B (experiencing), questions were asked during the trip. The eMotion application was tracking plausible traffic conditions and when they were met, the application prompted the participant to reflect on his/her current anger and frustration levels (Figure 5.10). For a congestion criterion, we set a period of 1 to 2 minutes with a driving speed below 5 km/h. Audible notifications were used to inform participants when a question appeared. Questions disappeared when more than 30 seconds had elapsed without receiving an answer. At this point, it is important to note that participants were specifically instructed to avoid answering if they felt that there was any safety risk involved. Additional questions were asked at the beginning and at the end of every commute across all conditions including a question on participants' mood using a five-point Likert scale (i.e., "*How do you feel today?*"), ranging from 1 (very bad) to 5 (very good). Finally, upon every arrival, participants were instructed to end the study session and answered one more question about any unexpected congestions occurring during the journey ("*Was there any congestion you did not expect before?*").

In summary, regarding research question (i) we found no significant difference between predicted and experienced anger and frustration. Furthermore, there was no significant difference between recalled and actual anger, implying that individuals can accurately recall their experienced anger of a past commuting episode. With regard to frustration, we unveiled that participants' memory is systematically biased, with participants recalling the experienced frustration as significantly lower than what they actually experienced during the traffic congestion. We attribute this phenomenon to the manifestation of the "rosy view" effect, as previously hypothesized (H1). We also found that time of day influences the prediction of anger, which is predicted significantly higher in the morning than in the evening (ii). Our results show that mood is related to the prediction of frustration. The better the mood, the less frustration is predicted (H2). These differences could be explained by the fact that anger is directed towards a specific person in a concrete situation [214], and thus less influenced by individual issues like mood.

5.4.2 Retrospective In-Car UX Evaluation

Over the last decade, cars have been increasingly fitted with "infotainment systems" that rival modern smartphones. Several commonly found features on smartphones have now become standard in-car functionalities: built-in navigation systems with speech recognition capabilities, reading out incoming texts and e-mail messages, social media connectivity, and augmented reality projections are only a few examples. This tendency of integrating, if not replicating smartphone functionality through in-car infotainment systems can be seen as the result of strict regulation against the use of mobile devices during driving, in combination with long commute times. In fact, US commuters spend on average 74.9 minutes driving daily [137], while there is a steadily increasing number of vehicles in roads. Undoubtedly, either commuting or travelling, people spend a significant amount of time in their cars. This creates opportunities for designing new products and services that improve the in-car user experience (UX). Thus, the automotive field has long since been influenced by HCI standards and guidelines [131]. In fact, several mechanical controls and electrical systems have now fully transformed to the digital world and become adjustable via touch displays or become fully automated (e.g. cruise control) [206]. However, with the rapid growth of Information and Communication Technologies (ICT) that aim at providing experiences to car passengers, such as technologies that reduce stress during commuting [251], new methods for measuring in-car experiences are required. In this section, we describe an updated version of eMotion (i.e., $eMotion^{+}$) that utilizes event-driven capture for producing memory cues (RQ1) that support retrospective in-car UX evaluation. $eMotion^{+}$ captures contextual information as memory cues for the later recall of one's in-car experiences, instead of utilizing ESM, with the aim to provide useful insights to car interface designers and urban planners. In the next subsections, we showcase the need for evaluation techniques that respect drivers' and passengers' safety, present our theoretical background, and illustrate the $eMotion^{+}$ prototype.

Measuring UX in Automotive

The ubiquity of mobile devices have lead to their broad utilization for acquiring user feedback in various contexts and often in large scale longitudinal studies [151]. One of the most popular methods for capturing experience in situ is the Experience Sampling Method [142], which can be easily applied via mobile devices for outreaching participants. As we have previously seen, ESM is often thought as the gold standard for in-situ measurement since it samples experiences and behaviours right at the moment of their occurrence and thus, reducing memory and social biases during self-reporting [208]. However, it comes with certain disadvantages such as interrupting users' current activity while demanding an additional reporting effort. In fact, in certain cases ESM might turn out to be inadequate or even dangerous to employ. Automotive is a field where ESM is rather innately limited for collecting user feedback, as it may distract one from the primary task of driving. On the other hand, Day Reconstruction Method (DRM) was proposed by Kahneman et al. as an alternative to ESM [124]. DRM uses a retrospective self-report protocol that imposes a chronological order when a participant recalls her experiences at the end of the day. DRM was found as a good approximation to ESM and is well adopted in the HCI community. In our prior work, we have shown that picture of one's facial expressions [165] as well as visual lifelogs obtained with smartphones or wearable cameras [164] can assist individuals in recalling their experiences and behaviours, approximating the validity of Experience Sampling data. Here, we propose the enhancement of the DRM review process at the end of a given day, with the provision of visual and contextual information in the form of memory cues. We expect that our approach will improve further the accuracy of one's recall of past momentary emotions and hence, provide a viable alternative to in-car UX evaluation.

Building Effective Memory Cues

As we mentioned before, today's mobile technology comes with constantly increasing capabilities in sensing and capturing contextual information. Such contextual information may vary from simple location coordinates and time, to acceleration and orientation values that can infer activity types (e.g., walking, running, or sitting, etc.). In fact, even the simplest interaction with a mobile phone can generate contextual information such as sending an SMS or making a call. This contextual information can be then used as **memory cues** — hints (stimuli) that help one recall a past experience. As we have seen, the most effective memory cues are the visual ones (i.e., pictures and video), however, when combined with additional cues from different sources, their efficiency in triggering memories may increase further. As such, location cues have been found to support recall by inferring patterns of behaviour if they differ significantly [243]. Time on the other hand is central to human memory and past episodes are placed in temporal order when registered in memory [226]. Thus, it is easier to recall temporal adjacent events that occurred in temporal order than in a random order. With $eMotion^+$, we attempt through various combinations of different memory cues (e.g., location, time, and speed) to enhance visual cues (i.e., video, and pictures) and ultimately improve one's ability to recall a past in-car experience.

$eMotion^+$

Leveraging on the cued-based augmented memory recall approach, the $eMotion^+$ application now collects contextual information throughout a drive, and utilizes it later as memory cues for amplifying the driver's memory about momentary emotions while driving, and thus retrospectively assessing automotive UX. Particularly, the front facing camera records 10-second segments of driver's facial expressions. The back facing camera captures a snapshot of the view that the driver has on the road. Additional information is captured such as current speed time and location. Capture can be set to occur continuously or based on some criteria. For example, when the driver surpasses a certain speed limit for certain time frames (e.g., $>$ 10 Km/h for 2 minutes) or when the driver approaches

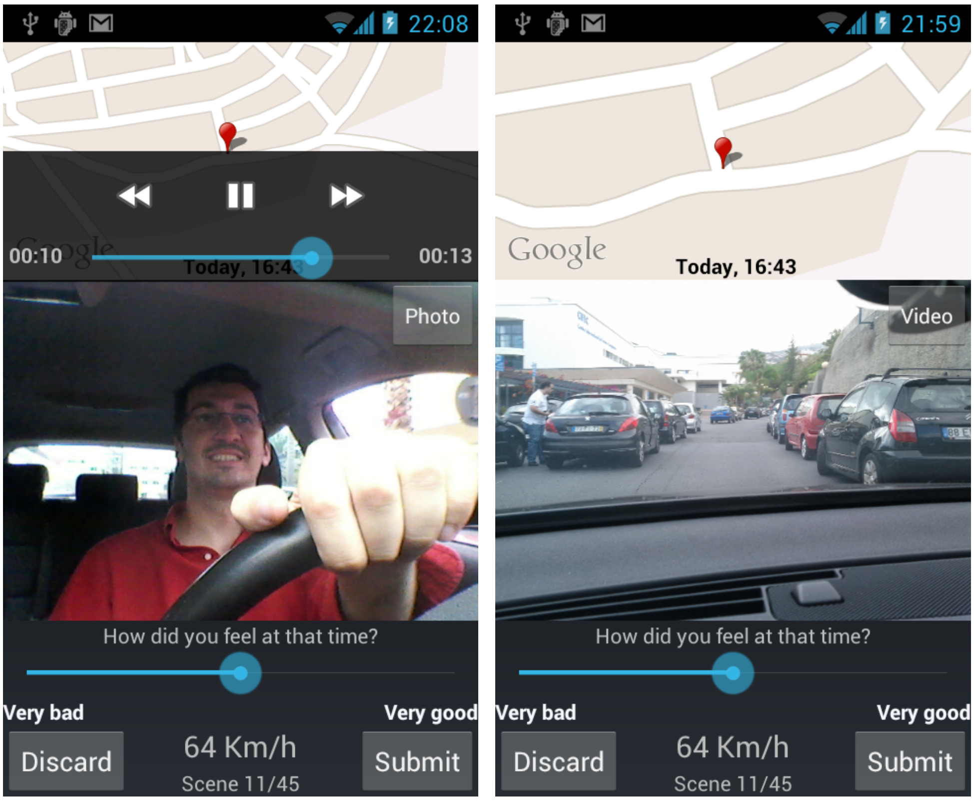


Figure 5.11. *eMotion*$^{+}$ combines four types of memory cues to assist users in recalling their experiences in particular moments of the trip: time, location, a 10-second "selfie"-video, and an image of the road.

certain locations or path (e.g., at the 2nd crossroad on highway A12). The mobile device can be mounted on the windscreen with a holder in a similar way one would use a mobile device for navigation purposes (see Figure 5.11). Extra care needs to be taken so that the back and front cameras are not occluded. In review mode, the $eMotion^{+}$ application utilizes all collected contextual information for forming memory cues that are presented together via a novel interface, as shown in Figure 5.11. On the top left the timestamp indicates the time a scene was recorded, the map shows the driver's location at that time, while 10-second video segment of driver's facial expressions is played continuously. A semi-transparent button allows a user to change between the in-car video and the road picture for acquiring additional memory cues. Speed at that time is displayed at the bottom. The user can use a continuous scale bar to rate particular feelings (e.g., stress and frustration during commuting) as one may infer from or recall with the assistance of these logs. Two buttons in the bottom left and bottom right are used for discarding and submitting a rating, and move on to a next scene. The combination of driver's facial expressions video with road picture, location, speed, and time aims at maximizing the users' ability to recall the specifics of that particular moment. A bar at the bottom part of the interface is used for answering the question asked each time in a Likert scale. After feedback is obtained, location, time, and ratings are uploaded to an on-line database. No pictures or videos are uploaded, thus protecting user's privacy.

5.5 Discussion

Existing methods for ecological momentary assessment, such as the Experience Sampling Method (ESM) and the Day Reconstruction Method (DRM), have somewhat complementary limitations: While Experience Sampling is often too intrusive for participants' daily lives due to its repetitive prompting, Day Reconstruction in turn often suffers from partial memory bias due to incomplete recollection of one's behaviours and experiences. In this chapter, we first illustrated EmoSnaps, a mobile application that captures self-face pictures using the front-facing camera of mobile devices, and uses these pictures to assist the later reconstruction of one's experienced emotions. EmoSnaps advances existing work on ESM and DRM; in that, capture is almost invisible to the user, while reconstruction is enhanced by the self-face pictures, so individuals do not rely merely on their memory.

In Chapter 4 and the current chapter, we reported two follow-up studies that investigated the validity of EmoSnaps in real life. The first study revealed that by increasing the temporal difference between capturing and recall of an experience, we increase users' ability to infer emotion from their self-face pictures. The significance of this finding needs to be noted as it suggests that designers, contrary to common sense, should avoid employing EmoSnaps or related approaches for recent experiences, but rather employ this to "recall" experiences that lie further in the past. In the second study, we inquired into the potential of EmoSnaps to capture the nuances of mobile usage and everyday life. By deploying EmoSnaps "in the wild", we were able to investigate a larger set of mobile interactions, confirming that participants exhibited a checking habit formation, where they frequently checked their mobile devices but without engaging in lengthy interactions. As shown in the literature, this checking habit formation provides an instant gratification that may lead to an overall increase of mobile phone usage [174]. For instance, we found established patterns of use with high mobile content consumption over the morning hours (8:00 to 9:00) and a significant increase in users' levels of happiness over these hours. This could potentially indicate low levels of happiness experienced during early wake up triggering mobile content consumption and gratification derived from this consumption, leading to the experience of positive emotions. Finally, we also found diurnal and weekly happiness patterns, as well as identified interaction types that entailed a higher degree of happiness than others, as derived from self-face pictures.

Next, we provided a systematic investigation of predicted and recalled experience in comparison to actual experience in the context of daily commute using eMotion, a custom ESM mobile application. To the best of our knowledge, no such study exists to date. With these insights provide a holistic view on commuters' everyday emotions and experiences — not only when being on the road, but also before and after the commute. Thus, we extend the design space for the automotive context. We are confident that the presented study is a first step towards investigating driving experiences from a comprehensive (temporal) perspective, providing valuable insights and inspiration for the development of future automotive systems.

Finally, leveraging on the innate shortcomings of ESM to safely assess driver's UX, and the effectiveness of EmoSnaps in amplifying emotion recall, we designed and developed the $eMotion^{+}$ mobile application. $eMotion^{+}$ differs from the original eMotion application in that it employs unobtrusive event-driven capture of contextual information utilized later as memory cues, for assisting the emotion recall process. Despite we did not have time to evaluate $eMotion^{+}$ in a user study with actual drivers, we expect that it will be able to approximate established self-reporting methods such as ESM, and even surpass alternative ones such as

DRM. We also expect that the display of dynamic visual cues (i.e., video and pictures) will particularly amplify one's ability to accurately recall a past drive. With $eMotion^+$, we wish to contribute to the field of in-car UX evaluation with an approach that respects the driver's and passengers' safety, is adequate for field deployments, and at the same time increases the external validity of the evaluation results.

5.6 Summary

Overall, the results suggest that episodic memory augmentation, via event-driven capture, can be a viable approach to established methods of ecological momentary assessment, such as ESM, for measuring UX in mobile context. Users were able to recall emotions based upon their self-face pictures, with considerable accuracy, even a week after the sampling process, while commenting on its transparency. Generally, participants reported that the task of emotionally assigning their self-face pictures was easy and that they believe they were improved in judging their emotions after some task repetitions. Therefore, it would be plausible to assume that results from the two EmoSnaps deployments lend credence to similar approaches that utilize an even larger and more dynamic set of cues (e.g., images and video). As such, we believe that $eMotion^+$ could be a useful tool that amplifies one's recall for assessing driver's UX safely and effectively. Future work could focus on exploring new mobile/vehicular interactions for capturing self-face pictures/video of a mobile user/driver, and investigate how this could augment the experience recall for mobile/vehicular application prototype evaluation purposes.

All in all, our findings highlight the potential of event-driven capture in augmenting episodic memory recall, with self-face pictures as memory cues, in practice (i.e., RQ1). Insights gained from various field deployments, contribute to the design and development of future pervasive memory augmentation systems (see Chapter 9 for a summary) that utilize cue-based augmented memory recall. In the next chapter, we attempt to better comprehend the influence of pictures on human memory, by systematically investigating how picture capture modality affects human memory and the quality of memory cues produced.

Chapter 6

Capture Modality Effect on Memory Recall

As we have seen in Chapter 2, today's abundance of cheap digital storage in the form of tiny memory cards put literally no boundaries on the number of images one can capture with one's digital camera or smartphone during an event. However, prior work has shown that taking many pictures may actually make us remember less of a particular event. Does automated picture taking (lifelogging) help avoid this, yet still offer to capture meaningful pictures? In this Chapter, we investigate the effect of capture modality (i.e., limited, unlimited, automatic, and no capture) on people's ability to recall a past event, and the quality of the memory cues produced through these modalities (RQ2). Our results from a field experiment with 83 participants show that capturing fewer pictures does not necessarily lead to the capture of more relevant pictures. However, when controlling for number of pictures taken, our results show that having a limited number of pictures to capture may lead to pictures with increased memory value. At the same time, automated capture failed to produce pictures that would help remember the past experience better.

6.1 Author's Contribution

The author of this thesis had a leading role in the study reported in this chapter. In particular, his contribution includes the conceptualization and design of the memory intervention, the study design, the data analyses, and writing. The co-authors of the original publication [164] conducted the user study and collected the data. Senior co-authors provided useful guidance and expertise on human memory theory as well as, editing the original publication. For more information, see the original publications [164, 167].

6.2 Introduction

Modern digital storage and sensor miniaturization (in particular digital cameras) have created entirely new forms of capture practices, such as "lifelogging" [209]. As we have seen, lifelogging entails the capture of personal experiences in an automated and continuous fashion, utilizing hardware that includes not only (small) digital cameras but also positioning technology and physiological sensors. Lifelogging offers support for a range of activities. At the outset, it helps to simply record experiences (e.g., trips and sports) for sharing with others, later reminiscence, or for security reasons (e.g., a car dash cam). It also allows people to keep an overview of their progress and habits (e.g., smoking or eating), and serves as a personal reflection tool. Davies et al. envision the next step of such captured data to be the augmentation of human memory [56]. Instead of simply providing a convenient way to look up information once it is needed, Davies et al. suggest to use the data to prepare special memory cues that, when played back to the user in an ambient fashion, will reinforce recall and thus make it easier to remember the captured experience.

While lifelogging cameras can easily capture a comprehensive account of one's daily activities, the challenge is to select the few key moments to play back to the user as memory cues. This prompts 2 questions: (1) *Are lifelogging cameras able to produce images that can serve as efficient memory cues?* And (2) *which images from a captured experience are best when used as memory cues?* We approach these questions with a simple experiment, which investigates the effect of three different modes of picture capture — limited, unlimited, and automatic — on the ability of people to recall a past experience with and without the support of the captured pictures. Hence, we formed in total four conditions: Limited, in which participants could capture a limited number of pictures, unlimited, in which participants could capture as many pictures as they wished, automatic, in which participants relied on a lifelogging camera to capture pictures, and finally no tech, in which participants took no pictures at all. All participants, in all three conditions that entailed picture capture were instructed to in the similar manner, so that they use any available picture capture means at their disposal for maximizing their recall capabilities about the experience at later stage. By contrasting manual with automated picture taking, we attempt to understand if (and how) conscious image framing results in images that are better able to act as memory cues later. By contrasting limited and unlimited picture capture, we try to understand if fewer images will lead to better memory cues. We ran a "campus tour" with 83 participants, grouped into different capture conditions, and asked them later to recall their experience with and without the use of pictures they had captured. Participants in the "automatic" condition were equipped with a 1st generation Narrative Clip[1] — they were able to focus on the tour and were able to draw on literally hundreds of images to remember the tour later. Those in the "unlimited" group were allowed to use the standard camera application in an Android phone to take as many pictures as they wanted — this should produce high quality images but may be affected by the "photo-taking impairment effect". The "photo-taking impairment effect" describes the detrimental impact that manual picture capture may entail on the formation of strong memories [108]. For limiting the number of pictures one can capture, we used a mobile application called "My Good Old Kodak" (MGOK) [167], which artificially limits the number of pictures one can take with a smartphone to 24 — we expected this to balance image quality with photo-taking impairment.

6.3 Striving for Selectivity

People capture content (e.g., video, pictures, etc.) manually and/or automatically (i.e., lifelogging), in an effort to amplify and prolong positive emotional experiences. However, as we have seen in Chapter 3, lifelogging and automatic capture come with significant drawbacks. The "always-on" capture fashion of lifelogging entails the capture of an enormous volume of (visual) information that is often irrelevant to the experience one strives to record and subsequently remember, thus leading experts to advocate for the need of "*selectivity, not total capture*" in lifelogging [210]. This selectivity is often considered as a necessity in designing successful and usable lifelogging systems [243]. In fact, a sizeable body of research is dedicated to filtering content in lifelogging, either pre- or post-capture, for limiting the content one has to review in order to remember. As a post–capture example, Ehlen et al. propose the utilization of machine learning techniques for creating a meeting summarization tool that performs natural language processing on participants' recorded utterances and subsequent topic analysis on the transcripts to summarize topics discussed [67]. Several pre-capture examples propose event driven sampling for selecting which moments are significant for assisting one's later recall and elicit User Experience (UX) quality levels. In Chapter 4, we introduced EmoSnaps, a mobile application that captures unobtrusively one's facial expressions based on a set of predefined actions on a

[1] http://getnarrative.com/narrative-clip-1

mobile device (e.g., slide in) and uses them for the later recall of one's momentary emotions [165]. Similarly, in Chapter 5, we illustrated $eMotion^{+}$, a mobile application prototype that captures unobtrusively short clips of drivers facial expressions along with road pictures during driving and additional contextual information under certain criteria (e.g., location and speed) and plays them back later to the driver for safely assessing drivers UX off-road [166]. Other lifelogging approaches introduce the use of physiological sensors hosted in a wristband for inferring a user's arousal levels and drive picture capture (pre capture) [168] or picture selection (post capture) [200]. In fact, Sas et al., found that pictures captured during higher GSR (Galvanic Skin Response) levels were able to trigger significantly richer recollections when reviewed, than pictures captured during lower GSR levels [200].

Figure 6.1. User Interface of "My Good Old Kodak" (MGOK) application, the Camera application replacement installed for participants in Group C (Limited). The screenshot shows the configuration we used for the campus tour, with a maximum number of 24 pictures.

A different approach, which puts the selection control back into the hands of the user, is our "My Good Old Kodak" (MGOK) mobile application [167]. The MGOK application (see Figure 6.1) attempts to reintroduce the classic paradigm of the old film cameras by artificially restricting the number of the pictures one can take, and at the same time hiding them from the devices photo gallery. The authors hypothesize that an imposed capture limitation would also result in taking pictures of higher importance, which would later support recall significantly better than ordinary pictures do. While Henkel [108] showed that the mere act of picture taking can be disruptive for the formation of new memories, and thus reduce the quality of recall (the so-called "photo-taking impairment effect"), Henkel's study did not have any limitations in the number of pictures taken. Also, recent work by Nightingale et al. [171] was unable to reproduce Henkel's findings.

In this chapter, we thus want to contrast three distinct types of capture modality — *limited*, *unlimited* and *automatic* — in terms of their impact on human memory (i.e., through the act of capture) as well as, their suitability for generating memory cues. Our contribution is hence threefold: First, we measurably estimate the *memory loss* that a capture modality imposes on those that take pictures during an event (i.e., quantify the photo-taking impairment effect across further capture modalities) [108]. Second, we measurably estimate the added value (i.e., *memory gain*) of pictures originating from all three capture types on one's ability to recall an experience at a later stage. Last but not least, we attempt to differentiate between two different explanations for the photo-taking impairment effect: (a) *due to the distraction caused by manual picture taking*, or (b) *due to disruption at encoding as a result of having external memory support* [212, 218]. Among others, we were also able to obtain the level of engagement with the campus experience and perceived quality of captured pictures, as influenced by the use of different capture modalities.

We expect that the "photo-taking impairment effect" will manifest for participants that captured pictures manually when they are asked to recall any details about the experience (i.e., campus tour) at a later stage, without the support of the pictures they took. However, when these participants review their manually captured pictures, they will be able to recall significantly more details than those that captured pictures automatically with a lifelogging camera. Particularly, we believe that participants that captured a limited number of pictures throughout the campus tour experience (i.e., limited condition) they will exhibit a higher memory gain when reviewing these pictures, as an effect of resource scarcity that perhaps forced them to capture more meaningful pictures. Similarly, we expect pictures that were captured in the "limited" condition will be rated as more meaningful and of better quality.

6.4 Study

For examining the aforementioned picture capture effects, we organized a campus tour event, offered to undergraduate students at the University of Essex's Colchester Campus in the UK. A campus tour offers a fun experience for participants involved, but also provides for a structured experience, allowing us to investigate the effects of different picture capture strategies on different participant groups in a relatively controlled setting. Following a between-subjects study design, we split registered participants into four distinct groups, of approximately equal size:

- **Group A – No tech:** This group was not equipped with any capture technology and was intended as the control group. Participants simply attended the campus tour by following the researcher while listening to her for each place they were visiting.
- **Group B – Unlimited (smartphone with native camera application):** This group was equipped with smartphones that had a native camera application installed. Participants could take as many pictures as they wished during the campus tour.
- **Group C – Limited (smartphone with MGOK application):** This group was also equipped with smartphones but the native camera application was replaced with the MGOK application, which restricts the number of pictures each participant could take to 24. Similar to Group B, participants of Group C could capture a picture with the MGOK any time they wished.
- **Group D – Automatic (Narrative Clip):** Lastly, a group of participants was equipped with the Narrative Clip, a lightweight wearable camera that one can clip onto one's clothes (typically at or below neck height) for capturing pictures automatically every 30 seconds.

We formed the following hypotheses:

(H1) Participant groups that captured pictures manually (Group B – unlimited and C – limited) will exhibit lower memory scores at a later unassisted recall of the campus tour experience than groups that capture pictures automatically (Group D – automatic) or did not capture pictures at all (Group A – no tech), due to the "photo-taking impairment effect" [108]. For the same reason, we also expect lower engagement with the tour in Group B and C.

(H2) Pictures captured with the smartphones native camera application (Group B) will be rated as more difficult to review due to their increased quantity. In fact, this hypothesis is in line with prior work that stresses the need for "selectivity, not total capture" in lifelogging, in that carefully selected content may be more valuable for one's memory recall than large volumes of (often indifferent and highly similar) content [210]. However, we expect that, since they are manually captured, these pictures still hold a considerable potential to improve participants' ability to recall the campus tour experience, displaying increased memory value and reported as more memory supportive than pictures taken with the Narrative Clip (Group D).

(H3) Pictures captured with the MGOK application (Group C) will hold higher "memory gain" as compared to pictures captured in all other conditions, both in terms of memory scores after picture review and as rated by the participants. The imposed picture capture limitation should lead to a more selective capture behaviour and hence increased (perceived) value [149].

(H4) Pictures captured with the Narrative Clip (Group D) will have a much lower perceived quality than manually captured pictures (Group B and C), due to its automated operation [41]. We also expect that participants of Group D will rate their pictures as less meaningful and will indicate lower feeling of ownership and engagement when reviewing them.

6.4.1 Participants

In total, we recruited 83 participants (55 were female) with an average age of 25 years (M = 25.301, SD = 8.849). All participants were Essex University Psychology undergraduates, with varying levels of familiarity with the campus whereabouts and its past. Participants were compensated with £18 for their participation (except for first year psychology students who were given 3 course credits for one of their modules). Participants were recruited through the Psychology departments recruitment website for participating into a memory study that includes a campus tour.

Figure 6.2. Narrative Clip 1, used for automated capture.

6.4.2 Procedure

Participants were first informed about the study and its purpose, and then asked to sign an informed consent form. We randomly assigned each of the 83 participants to one of the four groups described above, in order to achieve a balanced distribution (Group A: 22; Group B: 20; Group C: 21; Group D: 20). All participants were told that they would participate in a memory experiment about the University Campus, in which a researcher would take them on a roughly 60-minutes tour of the campus, visiting a total of 10 distinct locations. Participants in groups B and C (limited and unlimited capture) were instructed to use a provided capture device (an Android smartphone) to take pictures that would help them later to better remember the places they visited. However, no examples on how to capture the tour were given (e.g., when and how to capture a picture). Participants in group D were given a Narrative Clip (see Figure 6.2), and were asked to pin this to their shirts or jackets during the tour. We explained participants how the Clip would automatically take a picture every 30 seconds, and that they could later use these pictures to aid their memory. Note that while the Clip also supports manual picture taking (by quickly double tapping it), we did not tell participants about this feature, as the goal was to investigate the Clips automated operation only. None of the participants in group D had seen or used a Narrative Clip before, and hence no one used this feature during their tour. No other picture capturing equipment was permitted during the tour for all conditions — neither digital cameras nor personal smartphones (participants had to leave these in the researchers office at the beginning of the tour). Participants in group A had thus no capture device at all with them during the tour.

Each tour group had between 2–4 persons, of which all were undergoing the same condition, i.e., all participants in a tour were using the same capture modality throughout their campus visit (a between-groups design). At each of the 10 distinct locations, participants where shown 6 specific items (identical items for all groups), and were told an interesting "fact" about each specific item. For example, at the library (a distinct location), participants were shown a printer, a sculpture, a painting, the floor plan, lifts, and Japanese dolls, and for each they were told a fact; for example, for the dolls the fact was "*These were donated by a Japanese diplomat*". The fact that our campus tour did not only focus on the geographical setting and its physical artefacts, but also on more or less arcane info at various landmarks, ensures that even students more familiar with the campus would still have a memorization task.

After the campus tour had ended, participants were invited individually to the lab, where they were given a 10-minutes time to relax during which they could use their personal smartphones to access the internet or chat on-line. During that time, we collected any material that was captured either with the smartphones given to the participants (i.e., both the native camera application and the MGOK application), or with the Narrative Clips. Immediately afterwards, participants were asked to perform a memory test of the campus tour experience[2]. Participants were presented with all the names of the distinct locations they had visited during the campus tour in a random order and were asked to write down for each distinct location as many specific item names as they could remember. For each distinct location, participants had 30 seconds before moving to the next one. After participants were finished, we collected their responses and calculated later their individual memory scores based on how many specific items they could accurately recall.

A week later, participants returned to the lab to perform a "delayed recall task". Participants were instructed to write down their experience of the campus tour in 10 different episodes, each one corresponding to the 10 distinct locations they had visited a week before. Participants were asked to describe those locations with as many details and events they could recall as possible. For this task, we allowed a maximum duration of 10 minutes in total. Then, participants that belonged to groups that captured images during the campus tour (i.e., Groups B, C or D) were given a maximum of ten minutes to review the content they previously captured, either with the native camera application, the MGOK application (Figure 6.1), or automatically using a Narrative Clip (Figure 6.2) (depending on which group they were in), on a computer in the room. We allowed a maximum time of 10 minutes for the content review session, although most participants found 10 minutes more than sufficient. We also recorded their individual content review time. For group D, we purposefully did not let participants use the Narrative application — the bundled mobile software that comes with the Narrative Clip, offering an automatically curated selection of a persons daily pictures (called "moments"), similar to services such as Google Photos. Instead, they simply received all pictures captured during the tour (typically between 100 to 110) in a file folder on the computer, similar to Groups B and C. While this entailed more effort than simply reviewing the curated selection made by a tool like the Narrative application, participants were still able to look through the pictures in a few minutes. Also, we did not want to test the quality of the Narrative

[2]Note that for the purpose of a second experiment, completely unrelated to the purpose of this article, half the participants in each experimental group received retrieval practice on half the specific items from half the locations before they started the memory test. Our data reported in this article thus excludes the results on these (five) locations for which this half of our participants received additional retrieval practice.

application in selecting important pictures, in particular since the Narrative application selection might filter out certain pictures (e.g., blurry ones), yet these may still hold information that could trigger episodic memory recall. After the review (which only took place for participants from Groups B, C, and D), participants were asked to perform again the serial recall task. Participants who did not capture pictures (Group A) were instead given an 8-minutes interval with no specific instructions, after which they then simply repeated the serial recall task process. Same as before, participants were asked to cluster their memories about the campus tour in 10 episodes corresponding to each of the 10 distinct locations they had visited a week earlier. We allowed a maximum duration of 10 minutes for the second serial recall task.

Upon ending, participants whose groups entailed picture capture answered a series of questions in a 5-point Likert scale, ranging from 1 ("not at all") to 5 ("very much"), inquiring into the perceived ability to manage the captured images, the quality of the images, their feeling of ownership over the images, how meaningful they found the captured pictures, as well as their subjective engagement during the review of their pictures. All participants (also those from Group A) were asked to report on how engaged they felt during the campus tour, using the same 5-point Likert scale. We also collected qualitative insights by asking participants to describe their feelings about the campus tour and (for Groups B, C, and D) whether the provided capture device was helpful in engaging and/or remembering about their campus tour experience.

6.4.3 Apparatus

As outlined in the study description, the only group of participants that relied solely on their memory for recalling the campus tour experience was Group A, thus acting as a control group. Participants in groups B and C were each given a Samsung Galaxy S3 Mini smartphone running Android OS version 4.2.2 with an 8 Megapixel camera resolution. While participants in Group B were asked to simply use the native Samsung camera application, participants in Group C were given smartphones in which we had replaced the native camera application with our MGOK application [167]. The MGOK application offers a maximum of 24 pictures and informs the user about the remaining shots they can take. It also requires a long-press on the shutter button in order to first focus and then capture a picture. Captured pictures are deliberately hidden from the devices photo gallery. The MGOK application (see Figure 6.1) thus tries to approximate as much as possible the traditional film cameras that required one to focus before taking a picture and develop the film before one could see the result.

Finally, Group D used the Narrative Clip 1 (see Figure 6.2), which we clipped to participants' clothes (typically on the neck of a shirt, or lapel of a coat) for capturing pictures automatically every 30 seconds. The Narrative Clip 1 is equipped with a 5-megapixel camera sensor and has a battery life of approximately 2 days. While the Narrative Clip also supports explicit/manual capture by double tapping on it, we intentionally did not inform participants about this feature. As the Clip tags such manual captures in the captured pictures metadata, we were able to verify that none of our participants actually used this feature.

Measure	**Group A**	**Group B**	**Group C**	**Group D**
$score_{AfterTour}$(%)	67.341 (15.077)	59.555 (12.007)	59.082 (15.415)	63.648 (13.88)
$score_{AfterAWeek}$(%)	57.811 (17.552)	47.5 (16.226)	42.698 (17.043)	50.907 (18.219)
$score_{AfterReview}$(%)	63.417 (19.481)	65.018 (21.184)	55.801 (21.004)	54.259 (21.526)
number of pictures	0 (0)	76.55 (25.787)	22.714 (2.512)	105 (12.579)
review time (mm:ss)	00:00 (00:00)	04:17 (02:29)	02:07 (01:24)	02:35 (00:46)
memory loss (%)	9.528 (12.166)	12.055 (13.79)	16.384 (16.023)	12.74 (12.39)
memory gain (%)	5.606 (8.412)	17.518 (18.686)	13.102 (12.998)	3.351 (10.398)
S1 "image quantity" (1–5)	N/A	2.6 (1.095)	1.904 (.83)	2.55 (1.316)
S2 "memory aid" (1–5)	N/A	4.3 (.978)	4.19 (.928)	3.5 (1.147)
S3 "ownership" (1–5)	N/A	4.2 (1.105)	3.19 (1.749)	2.65 (1.496)
S4 "image quality" (1–5)	N/A	3.25 (1.292)	3.571 (1.164)	2.85 (1.225)
S5 "semantic value" (1–5)	N/A	3.85 (.933)	3.238 (1.22)	2.3 (.801)
S6 "review engagement" (1–5)	N/A	3.6 (.994)	3.095 (1.135)	3.4 (.94)
S7 "tour engagement" (1–5)	4.09 (.683)	4.15 (.745)	3.809 (.813)	4.05 (.686)

Table 6.1. Overview of all measures used in the study and their descriptive statistics (i.e., average and standard deviation in brackets).

6.4.4 Measures

For examining our hypotheses, we employed a series of quantitative and qualitative measures, both of objective and subjective fashion. First, we measured each participant's ability to accurately recall the campus tour experience in 3 distinct stages, resulting in 3 memory scores: a) right after the campus tour ($score_{AfterTour}$), b) a week after the campus tour ($score_{AfterAWeek}$), and c) a week after the tour and after reviewing their content ($score_{AfterReview}$). Memory scores measured the percentage of specific items recalled from those shown at each distinct location in the campus tour, ranging from 0 (no recall) to 100 (absolute recall) and were obtained using the category-cued recall method [2, 229], with "spatial location" as the category and "items in that location" as the exemplars. This enabled us to perform comparisons between different participant groups/conditions and at different stages, and even compute memory gain and loss rates in % between recall stages. Also, based on memory scores recorded at different stages, we were able to quantify the memory loss that participants experienced a week after the campus tour, and the memory gain after the picture review. As memory loss we define the difference in % between memory scores recorded right after the campus tour ($score_{AfterTour}$) and memory scores a week after the campus tour was completed ($score_{AfterAWeek}$). We consider memory loss an indicator of memory deterioration due to the manual capture of pictures (i.e., "photo-taking impairment" effect [108]), and/or due to relying on pictures captured as an external memory prosthesis [212, 218]. As memory gain we define the difference in % between memory scores recorded a week after the campus tour ($score_{AfterTour}$), and memory scores after picture review ($score_{AfterReview}$). We consider memory gain an indicator of the potential that pictures hold in assisting one to recall a past experience. Additional measures were "number of pictures captured" and "picture review time" per condition. Next, for comparably inquiring into the participants' views over the content they captured and then reviewed, we used 7 subjective measures in a 5-point Likert scale fashion. In particular, we asked participants to indicate how much they agree from 1 ("not at all") to 5 ("very much") in a 5-point Likert scale with the following 7 statements:

(S1) **Image quantity:** "*I felt the pictures were too many to manage.*"
(S2) **Memory aid:** "*I think the pictures helped me remember of the campus tour.*"
(S3) **Ownership:** "*I feel like the pictures belonged to me / were mine.*"
(S4) **Image quality:** "*I think the pictures were of good quality.*"
(S5) **Semantic value:** "*I think the pictures captured meaningful content.*"
(S6) **Review engagement:** "*I think reviewing my pictures was an engaging (i.e., exciting) task.*"
(S7) **Tour engagement:** "*I think the campus tour was an engaging experience.*"

For an overview of all measures used in the study along with their descriptive statistics, see Table 6.1. Finally, we also collected participants' comments about the campus tour, the picture capture and picture review experiences.

6.5 Results

In this section, we present our analyses results grouped around our previously stated hypotheses. As mentioned in section 6.4.2 above, for the purpose of an experiment unrelated to the aims of this work, half of our participants received retrieval training for items from half (i.e., five) of the visited locations. Our results thus report data from our 83 participants as follows: 43 participants memory performance from 10 unpractised locations, together with 40 participants memory performance from the 5 unpractised locations that were unaffected by this otherwise unrelated memory intervention.

6.5.1 Photo-Taking Impairment Effect

First, we investigated the effect of picture capture on memory scores across all four groups (i.e., conditions). Particularly, we examined the effect of condition type on participants' memory scores at three points in time: right after the campus tour was completed, one week later before reviewing the pictures, and one week later after reviewing the pictures taken during the tour. In order to differentiate between two different explanations of the photo-taking impairment [108] effect, we need to compare two different sets of scores: if the effect is due to the distraction caused by manual picture capture, we expect the scores of Groups A and D to be higher than in groups B and C. If the effect is due to disruption at encoding because participants rely on having external memory support, we should observe Group A's score to be significantly higher than any of the other three groups.

Levene's tests of homogeneity and Shapiro-Wilk tests of normality confirmed the assumptions of homogeneity of variance and normality, respectively, of the dependent variables $score_{AfterTour}$, $score_{AfterAWeek}$, and $score_{AfterReview}$, for the independent variable condition type. A separate one-way analysis of variance was performed for each set of scores at the three different intervals ($score_{AfterTour}$, $score_{AfterAWeek}$, and $score_{AfterReview}$) to compare the effects of condition type (the independent variable). The analyses displayed no significant main effect for $score_{AfterTour}$ ($F(3,79) = 1.588, p = .199, \eta_p^2 = 057$), a significant main effect for $score_{AfterAWeek}$ ($F(3,79) = 2.904, p < .05, \eta_p^2 = .099$), and no significant main effect on $score_{AfterReview}$ ($F(3,79) = 1.374, p = .257, \eta_p^2 = .05$). This indicates that participants' memory scores right after the campus tour and a week later after picture review did not differ significantly across all conditions (see Figure 6.3). However, participants' memory scores a week after the campus tour, but before picture review, displayed a significant variation across different conditions. In fact, post hoc tests using the Bonferroni correction revealed that participants in Group A (No tech: $M = 57.811$ %, SD = 17.552 %) were able to recall significantly more about the campus tour a week after it was completed, than participants in Group C (Limited: $M = 42.698$ %, SD = 17.043 %, $p < .05$) did (see Figure 6.3). However, no significant difference was found between Group A and participants in Groups B (Unlimited: $M = 47.5$ %, SD = 16.226 %, $p =$

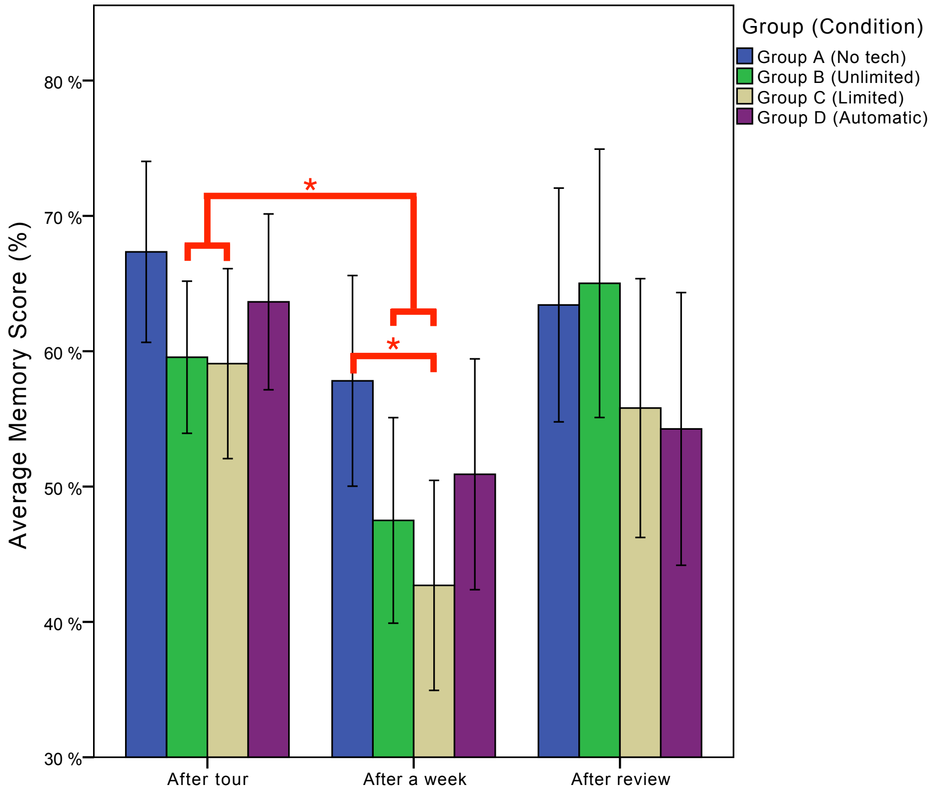


Figure 6.3. Participants' average memory scores (%) per condition: right after the campus tour, a week after and after picture review. Group A remembered significantly more than Group C a week after the campus tour. Statistically significant differences are marked with an asterisk for a value of $p < .05$.

.342), and D (Automatic: $M = 50.907$ %, SD = 18.219 %, $p = 1$), respectively. This seems to indicate that the photo-taking impairment effect is not simply due to having an external memory support (i.e., disruption at encoding). However, independent samples t-tests revealed significant differences in $score_{AfterTour}$ ($t(81) = -2.027, p < .05$) and $score_{AfterAWeek}$ ($t(81) = -2.493, p < .05$) between participants who captured pictures manually (Group B and C together) ($score_{AfterTour}$: $M = 59.313$, SD = 13.687 | $score_{AfterAWeek}$: $M = 45.04$, SD = 16.619) and those who did not (Group A and D together) ($score_{AfterTour}$: $M = 65.582$, SD = 14.463 | $score_{AfterAWeek}$: $M = 54.523$, SD = 17.994), indicating the presence of the "photo-taking impairment" effect due to the distraction caused by manual picture capture.

Next, we investigated if a possible presence of the "photo-taking impairment" effect would manifest as a lower self reported tour engagement (S7) for partic-

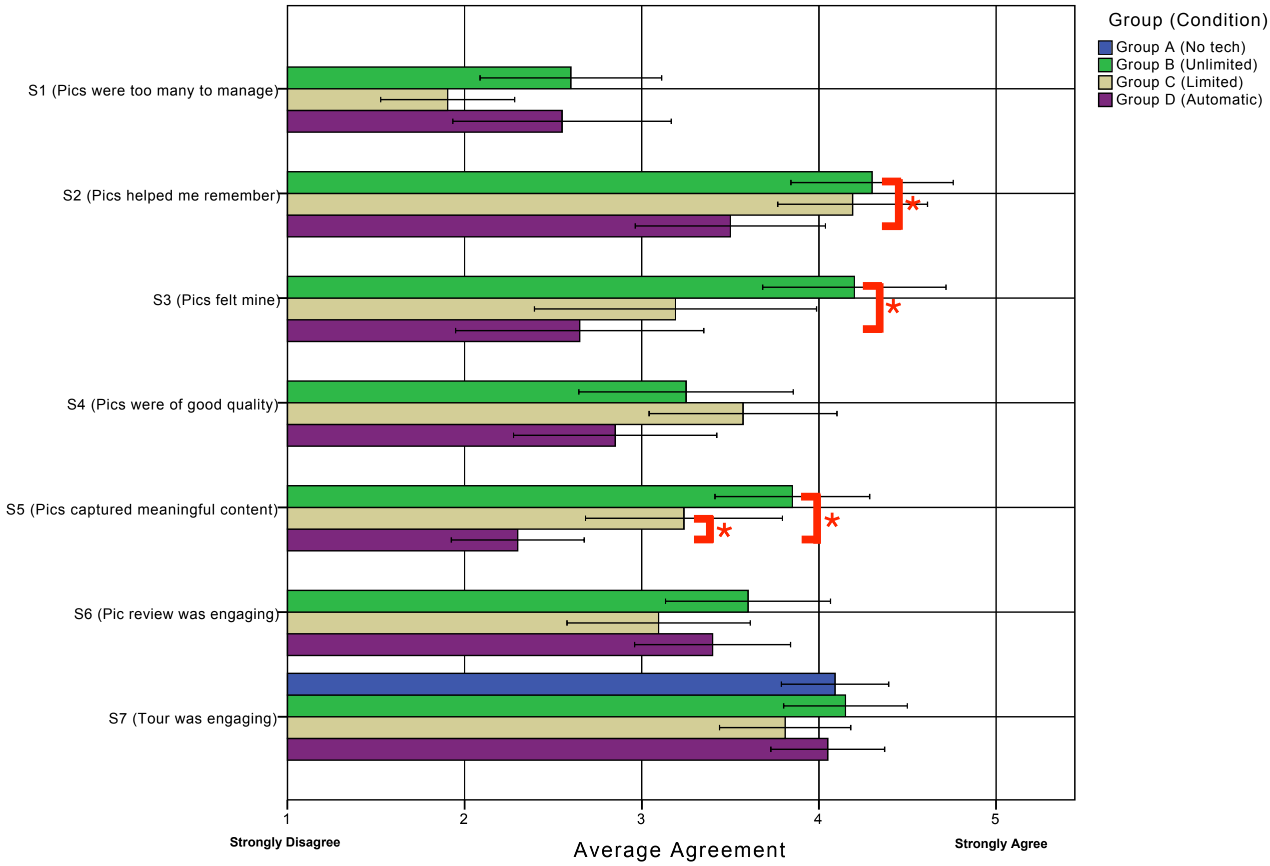


Figure 6.4. Average agreement with all statements (S1–S7) in scale from 1 ("Strongly Disagree") to 5 ("Strongly Agree") for the respective study conditions. We purposefully present averages (and not medians) for unveiling even the smaller differences across all conditions. Statistically significant differences are marked with an asterisk for a value of $p < .05$.

ipants who actively captured pictures during the tour (H1). In overall, Shapiro-Wilk tests of normality did not confirm the assumption of normality for the independent variable condition type for statements S1: "Image quantity" (B: $p = .033$, C: $p = .002$, D: $p = .032$), S2: "Memory aid" (B: $p = .000$, C: $p = .000$, D: $p = .007$), S3: "Ownership" (B: $p = .000$, C: $p = .000$, D: $p = .008$), S4: "Image quality" (B: $p = .028$, C: $p = .038$, D: $p = .128$), S5: "Semantic value" (B: $p = .01$, C: $p = .009$, D: $p = .012$), S6: "Review engagement" (B: $p = .005$, C: $p = .112$, D: $p = .018$), and S7: "Tour engagement" (B: $p = .001$, C: $p = .009$, D: $p = .001$). Non parametric Levene's tests of homogeneity did not confirm the assumption of homogeneity of variance for the independent variable condition type for statements S2: "Memory aid" ($p = .031$), S3: "Ownership" ($p = .013$) and S5: "Semantic value" ($p = .000$). Hence, for these statements, we proceeded with non parametric Moods Median Tests. For the remaining statements (S1: "Image quantity", S4: "Image quality", S6: "Review engagement" and S7: "Tour engagement"), non parametric Levene's tests of homogeneity confirmed the assumption of variance for the independent variable condition type, hence we proceeded with a series of non parametric Kruskal-Wallis tests. As such, for statement S7: "Tour engagement", a non-parametric Kruskal-Wallis with participants' tour engagement reported levels as dependent variable and condition type as independent variable displayed no significant effect for condition type (S7: $\chi^2(2) = 1.93, p = .381, \eta_p^2 = .032$). This again does not confirm H1, which assumes that participants who actively took pictures during the campus tour (Groups B/C) would report less engagement than those who did not (Groups A/D). A visual overview of the reported agreement with all statements can be found in Figure 6.4.

In general, participants perceived the campus tour as fun, engaging and at times educational experience.

> "[P43] *The campus tour was engaging I felt like a tourist...*"
> "[P9] *It is like trivial pursuit but quite interesting to learn some less known facts about the university.*"
> "[P2] *It was really exciting to know more about the campus.*"
> "[P10] *It was very nice to learn the facts about various things around campus.*"
> "[P30] *It was nice to go around places I knew and learn information about them as well as viewing sites I never did before... it made me feel more engaged with the campus.*"

Some participants got tired and distracted towards the end and felt that they were visiting places in an irrational order:

> "[P79] *The campus tour was educative but tiring.*"
> "[P1] *I started getting distracted by the end of it and paying less attention to the details...*"
> "[P47] *It starts to get a little boring when we visited a couple of places already.*"
> "[P22] *Campus tour was interesting, but the places were not arranged in accordance to distance to each other (it was more like a zigzag).*"
> "[P43] *The order of the locations was confusing because I normally do not visit these venues in that order.*"

Assuming participants' tiredness would influence their recall scores and be expressed in lower campus engagement (S7) levels, we computed Spearman rank-order correlation coefficients to assess the relationship between S7 and all memory scores. However, no significant correlation was found between S7 and all memory scores ($score_{AfterTour}$: $r_s = .018, p = .874, n = 83$ | $score_{AfterAWeek}$: $r_s = -.086, p = .441, n = 83$ | $score_{AfterReview}$: $r_s = -.049, p = .66, n = 83$).

6.5.2 Unlimited Manual Capture

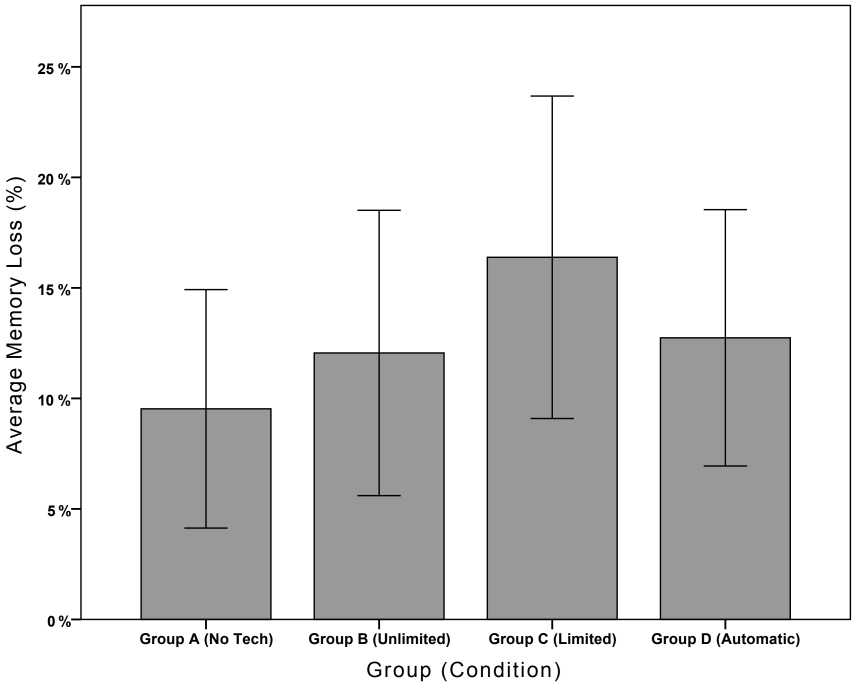


Figure 6.5. Average memory loss (%) per group/condition a week after the campus tour, without the support of captured pictures.

Next, we calculated the difference between memory scores achieved right after the campus tour was completed (i.e., $score_{AfterTour}$) and memory scores a week after the campus tour was completed (i.e., $score_{AfterAWeek}$), as a measure of memory deterioration and we encoded it in a new variable (i.e., memory loss in %). A one-way analysis of variance with participants' memory loss scores as a dependent variable, and condition type as an independent variable, displayed no significant main effect for condition type ($F(3, 79) = .917, p = .437, \eta_p^2 = .034$). This indicates that participants' memory loss rate over a week did not vary significantly across different picture capture modalities (see Figure 6.5).

Note that even though there was no significant difference in the memory loss rate, the overall memory scores (see Figure 6.3) still differed significantly (H1).

Similarly, for quantifying the added value of pictures in participants' ability to recall about the campus tour event that had taken place a week before the recall sessions, we created a memory gain variable (%). Memory gain is the difference between participants' memory scores before (i.e., $score_{AfterAWeek}$) and after (i.e., $score_{AfterReview}$) picture review, a week after the campus tour had been completed. A one-way analysis of variance with participants' memory gain rate as a dependent variable and condition type as an independent variable, displayed a significant main effect for condition type ($F(3,79) = 5.127, p < .05, \eta_p^2 = .163$).

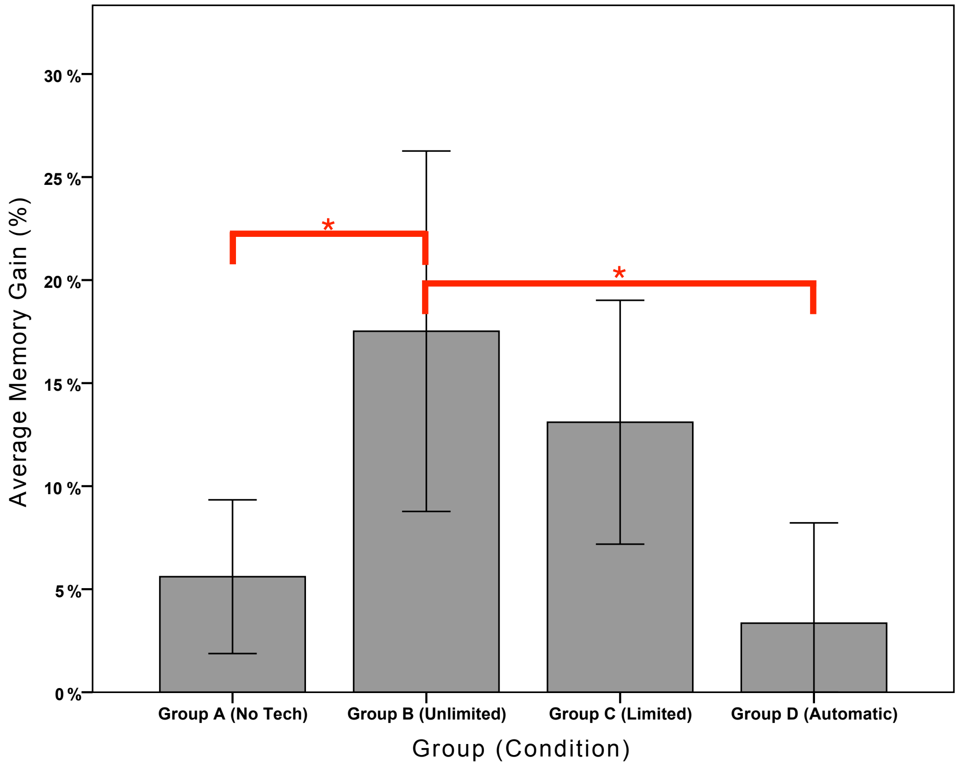


Figure 6.6. Average memory gain (%) per group/condition after picture review. Pictures captured with smartphones' native camera application (Group B) supported recall significantly more than pictures captured automatically with Narrative Clips (Group D). Group A did not review any content and simply repeated an unassisted recall session. Statistically significant differences are marked with an asterisk for a value of $p < .05$.

Post hoc tests using the Bonferroni correction revealed a significant difference in the memory gain rates of participants who reviewed pictures captured with smartphones (Unlimited: M = 17.518 %, SD = 18.686 %), as opposed to participants who reviewed pictures captured automatically with the Narrative Clip (Automatic: M = 3.351 %, SD = 10.398 %, $p < .05$), and participants who did not review pictures at all (No tech: M = 5.606 %, SD = 8.412 %, $p < .05$). In fact, a paired samples t-test with $score_{AfterAWeek}$ (M = 50.907 %, SD = 18.219 %) and $score_{AfterReview}$ (M = 54.259 %, SD = 21.526 %) for participants who used the Narrative Clip (i.e., Group D) revealed no significant memory improvement after the picture review ($t(19) = -1.442, p = .166$). This indicates that pictures captured with smartphones potentially hold a significantly higher memory gain than pictures taken with the Narrative Clip, or no pictures taken at all (see Figure 6.6). This finding is in line with hypothesis (H2) in that manually captured pictures hold a higher potential to increase a participants' ability to recall an experience.

Kruskal-Wallis tests with participants reported difficulty to review pictures (S1: "Image quantity") as dependent variable and condition type as independent, displayed no significant main effect for condition type (S1: $\chi^2(2) = 4.872, p = .088, \eta_p^2 = .082$) in contrast to an expected review difficulty imposed by a large number of pictures to review in the Group B (unlimited condition). However, a non parametric Moods Median Test with S2: "Memory aid" as dependent variable and condition type as independent variable, displayed a significant main effect for condition type ($\chi^2(2) = 7.174, p < .05$). Post hoc pairwise comparisons using Pearson's chi square tests revealed a significant difference in the medians between condition B (Mdn = 5, and D (Mdn = 4) ($\chi^2(1) = 7.03, p < .05, V = .419$), but no significant difference between condition C (Mdn = 4) and B (Mdn = 5) ($\chi^2(1) = .605, p = .437, V = .121$), or condition C ($Mdn$ = 4) and D (Mdn = 4) ($\chi^2(1) = 3.84, p = .05, V = .306$). This indicates that participants who reviewed pictures that they captured with smartphones (Group B) reported systematically higher memory aid than participants who reviewed pictures that they captured with the Narrative Clip. However, no significant difference was found between participants who reviewed pictures taken with MGOK (Group C), as opposed to participants who took pictures using smartphones (Group B) or the Narrative Clip (Group D) in terms of pictures reported memory aid. These findings corroborate H2 in that manually captured pictures hold higher memory value than automatically captured pictures.

Participants of both groups that entailed active picture capture (i.e., Groups B and C) expressed equally divided views over the disruptiveness of picture capture. Some found active picture capture contributing positively to their concentration and engagement:

> "[P4] *I think the photo capturing device helped me be more engaged during the campus tour in that as I was deciding how to take the picture, what to include...*"
> "[P6] *I was more concentrated on the information.*"
> "[P16] *...definitely feel that having the camera made me feel more engaged in the Campus tour.*"
> "[P46] *I feel that the capturing device did help me engage more during the campus tour.*"
> "[P57] *It definitely did help me to be more engaged because I knew I can't be too dependent on it as I had limited storage for pictures.*"
> "[P69] *I felt that being able to photograph places and items allowed me to better memorize some moments, as if I'd put a pin and the picture and leave it on a cork-board.*"

Others found active capture disruptive:

> "[P5] *The mobile phone given was in fact kind of restricting the tour.*"
> "[P7] *taking pictures made me feel like I was drawing too much attention.*"
> "[P52] *I don't think the device helped me in engaging with the campus tour. It has only interrupted me more as I have to listen to what the researcher said but at the same time I have to take the pictures too. I found it hard to focus...*"
> "[P70] *...spent too much time concentrating on taking a photo. Forgot or didn't hear facts because I was making sure I got a picture.*"
> "[P71] *Taking photos made me listen less to the actual facts.*"

Interestingly, some participants reported various capture strategies they developed or even changes in their behaviour that helped them deal with the specifics of each condition they were undergoing for maximizing memory gain later. For example, a participant in Group B (i.e., unlimited capture) reported increased awareness between what was captured and what was said for improving later recall:

> "[P77] *The camera made me more aware of how I would view the objects later in order to remember what was being said.*"

6.5.3 Limited Manual Capture Effect

One could attribute the significantly increased memory gain effect encountered in Group B to the unlimited number of pictures that participants could take. Ex-

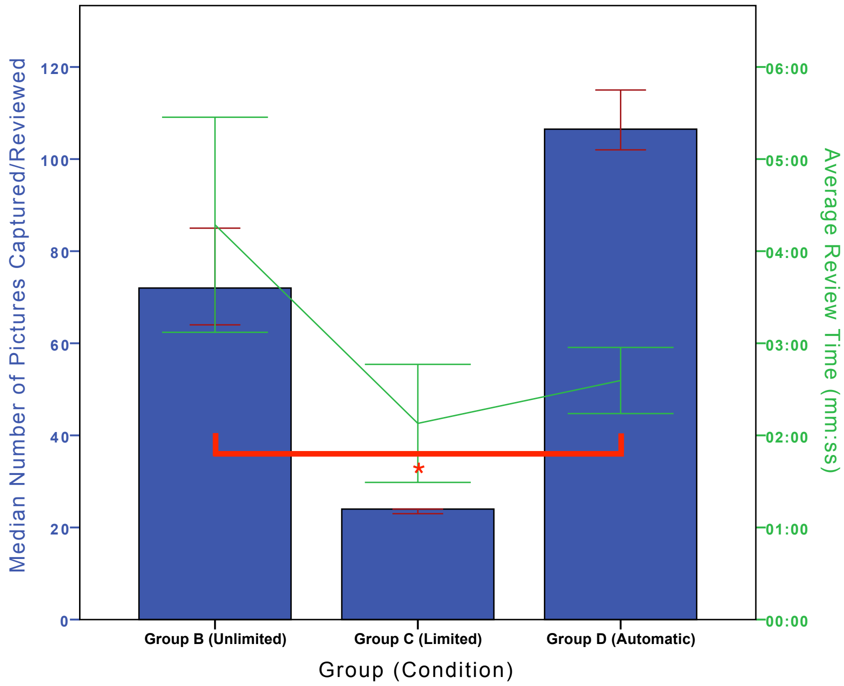


Figure 6.7. Left axis: Median picture number per condition. Right axis: Average picture review time per condition. Participants with Narrative clips (Group D) captured significantly higher number of pictures overall but also spent significantly less time reviewing them than participants using the native camera application (Group B). Median number of pictures is statistically different for all conditions. Statistically significant differences for average review time are marked with an asterisk for a value of $p < .05$.

pectedly, non parametric Moods Median tests displayed a significant main effect for condition type on number of pictures captured for conditions that entailed picture capture ($\chi^2(2) = 40.995, p < .001$). Post hoc pairwise comparisons using Pearson's chi square tests revealed that participants wearing the Narrative Clip (Automatic: $Mdn = 106.5$) captured a significantly higher number of pictures than participants who captured pictures with a smartphone (Unlimited: $Mdn = 72$) ($\chi^2(1) = 13.333, p < .001, V = .577$), and participants who captured pictures with the MGOK application (Limited C: $Mdn = 24$) ($\chi^2(1) = 41, p < .001, V = 1$) (see Figure 6.7). Moreover, participants using smartphones captured a significantly higher amount of pictures than participants who used the MGOK application ($\chi^2(1) = 13.887, p < .001, V = .582$) (see Figure 6.7).

Having found that "number of pictures" displayed systematic variations across all conditions, we next investigated its effects on memory loss and memory gain.

We first computed Pearson product-moment correlation coefficients for assessing the relationship between number of pictures taken and reviewed with memory loss and memory gain, respectively. However, we found no significant correlation between either number of pictures taken and memory loss ($r = .072, p = .519, n = 83$), or between number of pictures reviewed and memory gain ($r = .039, p = .727, n = 83$). Nevertheless, we suspected a plausible confounding effect of number of pictures on memory gain for condition type. For this, we performed an analysis of covariance [160] with memory gain as a dependent variable and condition type as an independent variable, while controlling for number of pictures captured/reviewed. The analysis maintained the assumption of homogeneity of regression ($F(2,58) = 2.563, p = .086$) and still displayed a significant main effect for condition type for the condition types that involved picture capture (B, C and D), after controlling for the effect of number of pictures variable ($F(2,57) = 5.402, p < .05, \eta_p^2 = .159$). Post hoc pairwise comparisons using the Bonferroni correction revealed significant differences for the adjusted averages between groups B (Unlimited: $M = 16.197$ %, SE = 3.375 %) and D (Automatic: $M = -2.098$ %, SE = 5.378, $p < .05$), but not between groups B and C (Limited: $M = 19.552$ %, SE = 5.99 %).

Despite influencing the effect of the condition type (i.e., B, C or D) on memory gain, the covariate number of pictures taken was not significantly related to memory gain per se ($F(1,57) = 1.595, p = .212, \eta_p^2 = .027$). This indicates that the number of pictures taken (and subsequently reviewed) has influenced the improvement in participants' ability to better recall the campus tour one week after, across all three groups that involved picture taking. When controlling for the number of pictures, Group C (i.e., limited) displayed the highest average memory gain ($M = 19.552$ %, SE = 5.99 %) surpassing the other two groups that involved picture capture (i.e., B: $M = 16.197$ %, SE = 3.375 % and D: $M = -2.098$ %, SE = 5.378 %), though not significantly (Figure 6.8). In principle, this finding appears to be in line with our hypothesis (H3) in that the MGOK application would produce pictures of higher memory value, though this trend did not emerge significantly.

During picture review, participants agreed in principle that pictures taken actively (Groups B and C) helped them recall details about the campus tour:

> "[P28] *Of course the photos helped me remember the campus tour.*"
> "[P44] *I found the pictures really helped me in recalling the items that I forgot. The pictures also help me to remember items associated to the places.*"
> "[P19] *The camera did help, as I remembered more items from the first*

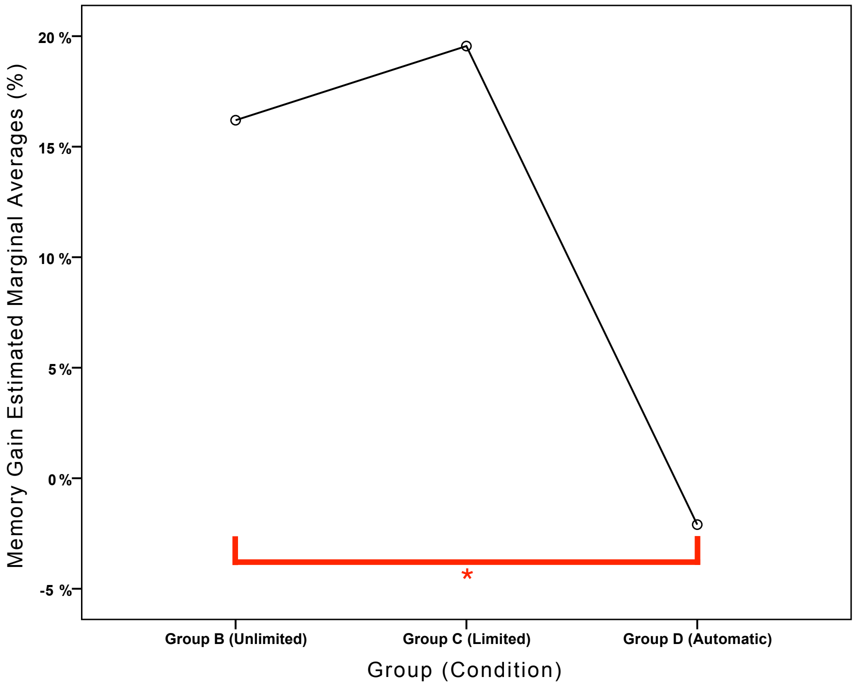


Figure 6.8. Estimated memory gain for all condition entailing picture review when controlling for number of pictures captured. Statistically significant differences are marked with an asterisk for a value of $p < .05$.

> *few locations that I photographed before they ran out.*"
> "[P69] *Some items were not very memorable. Having reviewed the item [picture], I can recollect where and when they were, but I did not spontaneously remember them.*"

Some also mentioned that people caught in the picture helped them remember more, as previously shown [145]:

> "[P37] *But the photo did help me remember the speaker...*"

In particular, the imposed capture limitation appears to have forced participants to develop some strategy for improving later recall. Some, as we hypothesized, strived to capture important places/moments during the tour:

> "[P53] *...by trying to take meaningful photos, which helped me to better remember the places and facts described.*"

Some used to the capture limitation to face the "recency effect" [229] by capturing more pictures in the beginning of the tour and less/none towards the end:

> "[P57] *I took pictures of everything in the beginning of the campus tour because I thought I would remember better of things that are recently shown and that I could refer to pictures that was taken earlier before.*"

Others regretted not having handled the limitation more efficiently:

> "[P72] *Yes, but I should have been careful because of the limited number of pictures available and the larger quantity of places to remember; I should have taken pictures of the most difficult to remember.*"

For some, picture capture with the MGOK application was perceived as a way to vividly imprint a scene in one's memory:

> "[P74] *I remember a lot of places in the same way I took a picture. It seems like I have those pictures in my mind. This was extremely exciting.*"

6.5.4 Unlimited Automatic Capture Effect

Although it was expected that the Narrative Clip would capture the highest number of pictures, we assumed that participants reviewing pictures captured via the Narrative Clip would also need significantly more time than all other participants. Nevertheless, an analysis of variance with picture review time in minutes as dependent variable and condition type as an independent variable displayed a significant main effect for condition type ($F(2, 58) = 8.963, p < .001, \eta_p^2 = .236$). Post hoc tests using the Bonferroni correction revealed that participants reviewing Narrative Clip pictures (Group D: M = 2:35, SD = 0:46) needed in average significantly less time in contrast to participants reviewing pictures taken with smartphones (Group B: M = 4:17, SD = 2:29, $p < .05$), despite having significantly more pictures to review in average (see Figure 6.4).

Moreover, non-parametric Moods Median Tests displayed a significant main effect both for condition type on S3: "Ownership" ($\chi^2(2) = 6.996, p < .05$) and S5: "Semantic value" ($\chi^2(2) = 18.523, p < .001$). For S3 "Ownership feeling", post hoc pairwise comparisons using Pearson's chi square tests revealed a significant difference in the medians between condition B (Mdn = 5) and D (Mdn = 2) ($\chi^2(1) = 7.03, p < .05, V = .419$), but no significant difference between condition C (Mdn = 4) and D (Mdn = 2) ($\chi^2(1) = 2.783, p = .095, V = .261$). This shows that during review, participants who took pictures with a smartphone reported a significant higher ownership feeling over their pictures, as opposed to participants who reviewed pictures taken automatically with the Narrative

Clip (H4). While a similar effect is observed between participants who reviewed pictures taken with MGOK application and participants who reviewed pictures captured with a Narrative clip, the difference is not significant.

For S5 "Semantic value", post hoc pairwise comparisons using Pearsons chi square tests revealed a significant difference in the medians between condition B ($Mdn = 4$) and D ($Mdn = 2$) ($\chi^2(1) = 18.027, p < .001, V = .671$) and condition C ($Mdn = 4$) and D ($Mdn = 2$) ($\chi^2(1) = 11.109, p < .05, V = .521$). This indicates that both participants who reviewed pictures captured with a smartphone and participants who reviewed pictures captured with the MGOK application, reported significantly higher semantic gain for these pictures, in contrast to participants who reviewed pictures taken with the Narrative Clip. This finding is aligned with H4 in that pictures captured manually (Group B and C) will be thought as more meaningful than pictures captured automatically (Group D).

In fact, participants of Group D (Narrative) did not favour the pictures to the same extent as those of Groups B and C:

> "[P22] *...but when I saw photos, they are not very meaningful.*"
> "[P23] *I don't think it helped me in remembering the things we saw, or in understanding the stories we heard better.*"
> "[P33] *Some of the pictures were helpful and some not, about half and half.*"
> "[P65] *Pictures provided doesn't help on the things that we need to remember.*"

Instead, pictures captured with the Narrative Clip helped participants recall better the temporal order in which the places were visited:

> "[P49] *I thought that later I would be able to look at the photos and instantly remember every single detail. It turned out differently, I simply reviewed the order of locations, but I could not remember everything.*"
> "[P20] *Seeing the pictures helped aid my memory in terms of the order in which the places were visited, and also helped me remember the places I had visited.*"

For the remaining statements (S4: "Image quality" and S6: "Review engagement"), series of non-parametric Kruskal-Wallis tests with S4 and S6 as dependent variables and condition type as independent variable displayed no significant main effect on S4 ($\chi^2(2) = 3.566, p = .168, \eta_p^2 = .059$) and S6 ($\chi^2(2) = 2.154, p = .341, \eta_p^2 = .035$) for the independent variable condition type. These findings did not confirm H4 in that pictures captured with the Narrative Clip will be rated as of significantly lower quality when compared with pictures captured with smartphones (Group B and C) and in that participants will be less engaged during reviewing Narrative Clip pictures (Group D).

Some participants mentioned positioning oneself in a way that helps capture the best picture with the Narrative Clip:

> "[P20] *I suppose I tried to position myself so that I took meaningful pictures, however, I could not be entirely sure I was managing to achieve this because I could not focus it like a camera.*"

Others reported wearing the Narrative Clip was something that they even forgot about:

> "[P11] *The capturing device was of secondary importance.*"
> "[P30] *I felt indifferent however about the camera and would perform the same in the experiment without it.*"
> "[P31] *I did not behave differently to what I would have done if I was not carrying a device.*"
> "[P33] *I forgot I was wearing the camera.*"
> "[P81] *I didn't put attention on the device, most of the time I don't even notice it exists.*"

Others mentioned that they became so reliant to the Narrative Clip that felt like carrying a "secondary brain":

> "[P49] *I think that the capturing device made me feel more confident in memorizing the information, because I thought that later I would be able to look at the photos and instantly remember every single detail... It helped me in being more engaged in the sense that it was like a second brain, i.e., if I missed something, the camera would surely capture it.*"
> "[P59] *I felt that I became reliant on the photo capturing device to help me with remembering what happened on the tour.*"

While some users blamed themselves for the lack of meaningful content captured with the Narrative Clip:

> "[P23] *In regard with the capturing device, I felt that I should be careful as I was taking photos automatically. When I saw the photos, they are not very meaningful, in fact I did not have an experience to manage to take better photos with the equipment.*"

6.6 Discussion

Below, we discuss our findings grouped by capture and review perspectives, for explaining how each condition affected participants' ability to recall the campus tour.

6.6.1 Memory Loss during Capture

We found that participants that used the MGOK application for capturing pictures during the campus tour displayed significantly greater memory loss a week later than participants who did not use any capture technology at all. However, this effect was far less pronounced for participants in either Group B (Unlimited) or Group D (Automatic). The introduction of a novel artefact (i.e., the MGOK application) with new features (e.g., capture limitation and focus to capture) may have distracted the participants more than those in Group B (Unlimited) who used a typical camera application throughout the campus tour. We were thus unable to confirm that the photo-taking impairment effect is due to disruption at encoding (i.e., participants relied on having external memory support so did not pay as much attention). However, when comparing Groups B&C (manual picture capture) with Groups A&D (no manual picture capture), we found significant differences both right after the tour and one week after the tour, lending credence to the explanation that the "photo-taking impairment" effect is due to the distraction at memory encoding caused by manual picture capture [108]. Interestingly, Group D (Automatic) did not display significantly lower memory loss when compared with Group C, as one would expect due to the unobtrusive capture fashion of the Narrative Clip. Some Group D participants reported that they "tried to position" themselves so that they could take "meaningful pictures", indicating that they were similarly distracted during the tour as those taking pictures manually. As for the difference between Groups B&C, the findings showed that manual capture both in limited and unlimited conditions was equally intrusive. However, participants in Groups B and C did not report a lower engagement than those who did not actively take pictures (Groups A – No tech & D – Automatic).

6.6.2 Memory Gain during Review

We found that pictures captured with the native camera application (i.e., Group B – Unlimited) offered significantly higher memory gain than pictures taken with the Narrative Clip. The same does not hold for pictures captured with the MGOK application, as opposed to pictures captured with Narrative Clip, or with the native camera application. However, when we controlled for the number of pictures captured (and subsequently reviewed), the analysis displayed higher memory gain for the pictures captured with the MGOK application, though not significantly higher. Moreover, participants rated their feeling of ownership and semantic gain higher for MGOK pictures than they did for Narrative Clip pictures. The higher memory gain shown for MGOK may be due to the imposed capture limitation that possibly led participants to capture more important moments, as we had hypothesized. However, further investigation is needed to reliably confirm this claim.

We discovered that reviewing Narrative Clip pictures did not provide any significant memory aid for the participants in Group D. Narrative Clip pictures were reported as holding systematically less memory aid, less ownership, and less semantic gain than pictures captured with the native camera application. In addition, while participants in Group D (Automatic) captured significantly higher number of pictures in comparison to all other conditions, they also took significantly less time to review them than participants who manually captured pictures with the native camera application (Group B – Unlimited). In fact, participants in Group D (Automatic) spent in average only 28 seconds longer for reviewing in average the quadruple number of pictures that participants in Group C (Limited) reviewed. These findings showcase that periodic automatic capture falls short in producing pictures that can effectively assist remembering. Furthermore, participants were in general disappointed with the pictures captured by the Narrative Clip and even accused themselves at times for not being experienced or not operating it appropriately. As reported, they used the pictures for eliciting the temporal order [48] of the places visited and thus, recall any details about the campus tour. Interestingly though, no significant differences were found on perceived image quality and quantity across all conditions, as we assumed beforehand.

6.6.3 Study Limitations

Participants in the study knew that the purpose of the experiment was to test memory, and therefore their aim was to utilize any means (if any) at their disposal for maximizing their recall at a later stage. While this is a typical experimental setting in cognitive psychology research, we usually do not aim to "maximize recall" in our day to day activities, hence one may question the ecological validity of the results. However, conscious attempts to remember our activities are still prevalent in our daily lives, from memorable events such as birthdays or reunions, to recreational activities such as museum visits, to busy days in the office. We also acknowledge that our study focused only on one particular aspect of lifelogging — the recollection of past experiences (episodic memory). However, as Sellen and Whittaker [210] described, lifelogging systems may have a wider range of benefits, which they summarized as the "Five Rs" (see Chapter 2): *Recollecting*, *Reminiscing*, *Retrieving*, *Reflecting*, and *Remembering Intentions*. Our study does not provide insights on how well automated capture systems such as the Narrative Clip could support these other benefits.

During the various campus tours we administered, we noted several times that participants in Group C — with the MGOK camera application — struggled with the unusual "hold-to-focus; release-to-shoot" shutter button functionality. This may have negatively influenced their memory scores as the act of taking a photo was more distractive than using a regular camera application. Equally influential might have been the choice of allowing only 24 photos — a slightly larger number (e.g., 36) might have still challenged participants to be selective in their picture-taking, yet supported a broader set of images. We also did not control for participants' experience in taking pictures with a mobile phone in general — less experienced participants might have equally been unable to follow the information offered by the tour guide while taking a photo. However, our participants' age ($M = 25.301$, SD $= 8.849$), as well as the fact that all of them owned a smartphone, suggests that all of them were reasonably familiar with smartphone image taking. Finally, we also did not control for participants' prior knowledge of the campus, which may have favoured some to be better able to remember individual items. However, by adding seldom known facts about each of the presented items, we expect that even those familiar with the campus had a wealth of new information to retain.

6.6.4 Implications and Future Work

We believe that our work can contribute to the design of future pervasive memory augmentation systems with respect to the idea of using wearable cameras as a replacement for manual picture taking. The "photo-taking impairment effect" would posit that wearable cameras offer less memory distraction, hence improve active memorization of events. The artificial limit on the number of photos taken offered another point in the design space, questioning if fewer, more "thoughtfully" taken pictures may lower this memory impairment. Our findings suggest that the quality of current generation wearable cameras does not yet live up to this promise, and that we require novel ways of capturing meaningful memories in a non-distractive fashion. Also, the fact that the wearer is not included in the captured pictures may limit the value of such images for recalling episodic memories. Bexheti et al., recently proposed a system architecture for automating the sharing of lifelogging images for co-located peers [24].

In future work, we plan to continue our trials towards investigating the added value of lifelogging images for human memory in later stages of recall (i.e., a month or a year after an experience). We believe that lifelogging cameras hold a potential for augmenting one's memory recall under certain conditions. For example, the need for selectivity and not total capture is one direction we are currently investigating [210]. In particular, we are investigating if a range of physiological responses (e.g., heart rate) as measured by wearable sensors (e.g., a smart watch) could indicate moments of increased significance or increased memory value for informing the capture or display of specific lifelogging content [168, 200].

6.7 Summary

In this chapter, we contrasted *limited*, *unlimited*, and *automatic* picture capture in augmenting human memory recall, and in the quality of memory cues produced (i.e., RQ2). We could confirm that manual picture capture may lead to the encoding of memories of lower quality. Contrary to Nightingale et al. [171], we thus confirmed Henkel's "photo-taking impairment" effect [108], and attributed it to the act of picture capture, not external memory support [212, 218]. We also found that automated capture as offered by today's wearable lifelogging cameras produces pictures that hold only a low potential in improving one's ability to remember a prior experience. While our participants exhibited various behaviours and techniques in an effort to handle an imposed capture scarcity (when using the "My Good Old Kodak" Camera application), limited capture did not improve their recall significantly, while unlimited capture did increase it significantly.

Findings reported in this chapter contribute to the design and development of future pervasive memory augmentation systems, as described in Chapter 9. In the next chapter, we investigate the potential of hybrid memory cues to synergistically support both episodic and semantic memory recall, while trialling the feasibility of cue-based memory augmentation in the workplace domain.

Chapter 7

Augmenting Memory Recall for Work Meetings

Being able to better recall a work meeting could improve coordination and collaboration among peers, ultimately raising overall productivity. For evaluating the effectiveness and feasibility of hybrid memory cues that target both episodic and semantic memory (i.e., RQ3), we selected the workplace application domain, and we conducted a multi–week study with seven groups. Using a within-subjects design, participants in the experimental condition were, prior to a meeting, briefly presented with an automatically created memory augmentation aid (slides), based on captured data from a prior meeting. Our results show that a 3–4 minute exposure to our simple image-keywords slide deck prior to a meeting, increased our participants' ability to recall their previous meeting by up to 15 %. Our findings serve as an initial baseline against which future memory augmentation systems can be compared.

7.1 Author's Contribution

The author of this thesis had a leading role in the study reported in this chapter. In particular, his contribution includes the conceptualization and design of the memory intervention, the study design, data analyses, deployment and writing. The co-authors of the original publication [169] assisted in conducting the user study and collecting the data. Moreover, the co-authors also implemented a prototype version of the envisioned system. Senior co-authors provided useful guidance and expertise on human memory theory as well as, editing the original publication. For more information, see [169].

7.2 Introduction

Work meetings present an ideal testbed for trialling the combinatory power of hybrid episodic-semantic memory cues (i.e., RQ3), since people generally strive for remembering work meetings better, yet are often unwilling to take extensive notes and/or review detailed minutes before subsequent meetings. We set up an experiment that enabled us to assess to which extent a pervasive memory augmentation system could improve one's recollection of past meetings. Our system simply creates a set of **memory cues** in the form of a short slide deck, extracted from data that it automatically captured during a previous meeting. Participants reviewed these slides prior to their next meeting to refresh their memory about the previous meeting. Using a technique known as a "*cognitive interview*" [78], we then determined the past meeting recollection of our participants both in a control condition (no memory cues shown), and in the experimental condition (after having browsed the memory cue slide deck). By measuring the topical overlap between a recorded meeting and each participant's probed recollection of the meeting, we were then able to *quantify* the effect that our memory augmentation system had on our participant's recollection of a prior meeting.

Obviously, reviewing any sort of summary of a past meeting prior to the next meeting *will* increase one's immediate recollection of said prior meeting. We would also expect a similar effect if one would review meeting notes, or watch a recording of a prior meeting. What is not clear, however, is the *extent* to which such a simple, automated memory aid can augment memory recall. To our knowledge, our study is the first that *quantifies* the actual increase in memory recall stemming from an automated memory augmentation system, and hence our findings contribute to the design and development of future pervasive memory augmentation systems.

7.3 Supporting Memory in the Workplace

The idea about a system that stores one's digital records (e.g., documents, images, multimedia, etc.) for a lifetime goes back to the 1945 vision of the *Memex* by Vannevar Bush. While Bush did not specify the exact technology for implementing his vision, he predicted an era when storage will be virtually unlimited. Fast forward to 2002, where the *MyLifeBits* project attempted to fulfill the promise of Bush's vision [84]. *MyLifeBits* started as a platform that could log all personal information generated and accessed on a PC, but its memory enhancing aspects quickly emerged [83]. However, the potential of technology to enhance work meetings was discussed long before *MyLifeBits* project was realized.

Project Nick theorized that the introduction of decision support systems (built on capture and presentation technology) in meetings can greatly facilitate productivity, given that meeting dynamics are well understood [49]. Yet, most research that explores the use of pervasive technology for work meetings focuses on increasing productivity and fostering collaboration, with only a few examples showcasing the importance of memory augmentation in meetings. Jaimes et al., present a technique that helps meeting attendees retrieve segments from video recordings of meetings, based on the use of some high-level search attributes (which they call *memory cues*) [119]. They support a finite set of these attributes (e.g., time when meeting happened, meeting room layout, meeting participants, use of equipment in a meeting), which they identified in a previous user study as most memorable to meeting participants. Note that the authors' concept of a memory cue as *something that you remember* is very different from our use of memory cues as *something that reminds you* (sometimes also called *retrieval cue*). Similarly, *MemTable* is an interactive table-top system for capturing and reviewing meeting group discussions [115]. It features several input modalities (e.g., simultaneous drawing, text entry, audio recording, image capture) to support heterogeneous collaboration styles. Users can then review such input from previous meetings (which are projected on a table-top's surface) by searching for items using text entry or by browsing prior content with the help of a time line. While both systems can help participants to better recall past meetings, they do so by means of an active search over captured data. In contrast, our goal is to understand whether we can use captured data of meetings to improve human memory, so that participants can recall details about a previous meeting without an external tool.

As we have previously seen (e.g., in Chapter 3), memory cues can come in many different forms and are per definition never *complete* — they do not comprise the entire experience but merely a related context, for example, a visual capture (picture), a sound (a song), or a smell [60, 165, 231]. A memory cue can also be textual, thus facilitating semantic recall, or an abstract visual that hints at one's activity, location, or social interaction at a certain time (e.g., device and application usage, GPS logs, Facebook posts) — a *personal context* of an episodic memory.

7.4 System

We adopted the visionary idea from [56] and built a memory augmentation system for supporting cued memory recall in the context of work meetings (see Figure 7.1 for an overview of the underlying process). Our system is comprised of (1) *an apparatus* for recording meetings in the form of audio-visual data, but also participants' physiological responses; and (2) *a component* for processing such captured logs in order to generate meeting memory cues.

In this section, we provide a short summary of both the capture system and memory cue generation component that we used in this study.

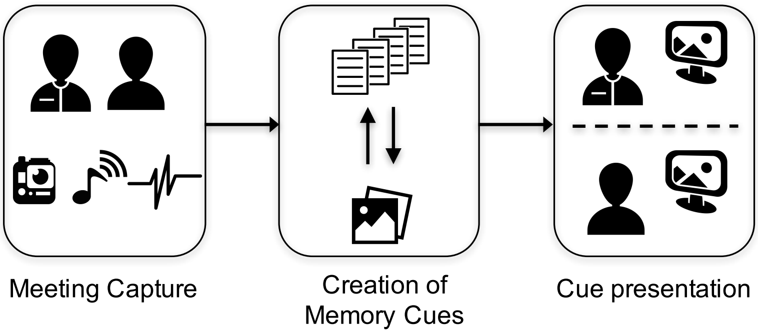


Figure 7.1. Overview of the memory augmentation process.

7.4.1 Experimental Apparatus

Our meeting capture equipment consists of: a *Narrative Clip 2*[1] wearable camera for capturing first-person view pictures (Figure 3.3 left); an *Empatica E4*[2] wristband for recording participants' physiological responses (Figure 3.3 right); and a *GoPro HERO4*[3] camera equipped with an external microphone for better quality audio capture of the meeting (Figure 3.4). The video camera served mainly as an audio recorder — the video material was simply captured in order to potentially help with speaker disambiguation during the transcription process. The reason for using a video camera instead of a simple audio recorder is the challenge of speaker disambiguation: after a small pre-study, we realized that a video recording offers additional visual cues for speaker identification, making audio transcription much easier.

[1] http://getnarrative.com
[2] http://empatica.com
[3] http://gopro.com

In this study, we did not use the double-tapping feature of the Narrative Clip 2, and thus did not explain this functionality to our participants (none of which had used the Clip previously). While we captured physiological data for all our participants during the meetings with the Empatica E4, we are planning to use it in another, yet-to-be-performed experiment, and thus do not report any findings related to E4 data here. During video recording with the GoPro HERO4 camera, we provided participants with a corresponding remote control to use as a *privacy button*, temporarily stopping the recording in case they wished to discuss something in private. After each recorded session, we downloaded all captured data and stored it in a secure repository hosted at our University. As each individual device required a slightly different download procedure (i.e., running a dedicated software for each different device), this step was performed manually by the researchers.

7.4.2 Creating Memory Cues for Meetings

The memory cues that are produced by our system are a combination of both Narrative Clip pictures and a set of keywords (*topics*) extracted from the audio captured around the time when each picture was taken. At the outset, the video recordings (captured by the GoPro) are transcribed to text in order to generate comprehensive meeting minutes, producing a time-stamped transcript with speaker identification. While this step can in principle be handled by automated transcription software, several commercial systems that we trialled were unable to produce sufficiently error-free transcriptions. We hence opted to manually transcribe the recorded videos (using a professional third-party service) in order to ensure that our experiment was not affected by the quality of the transcription system. Any final system would of course use a fully automated transcription process. The meeting transcripts are processed for extracting the salient topics using "SILAS", an analysis system for extracting topics from conversations [14]. In the rest of this section, we provide a brief description of the cue generation process; for a more detailed description of how SILAS works see [14].

In a first step, SILAS chunks the entire transcript Φ into smaller segments σ_i, using the "TextTiling" algorithm [106]. TextTiling divides a conversation into multi-paragraphs that represent passages or subtopics. Each segment σ_i is composed of a set of words w_i and a timestamp t_i (that indicates the moment in the conversation when that segment has started) denoted as: $\sigma_i \longleftarrow \{t_i, w_i\}$. Depending on the length of the meeting, our system produces in average about 20 such segments (as few as 4, as many as 38 in our data).

In a second step, the system extracts the most frequently spoken topics for each segment. For this purpose, SILAS uses the "Latent Dirichlet Allocation (LDA)" algorithm [25], a well-known technique in the domain of *topic modeling*. LDA is a probabilistic model that identifies the topics τ_i of a text corpus appearing in a document, where each topic is a probability distribution over words in the document. In our setup, we run LDA over the entire transcript Φ and for each segment σ_i it returns the set of topics that have the highest probability of appearing in that segment. This can be described as: $\tau_i \longleftarrow LDA(k, \Phi, w_i)$, where k indicates the total number of topics to be extracted, Φ is the entire set of words from the transcript and w_i is the set of words in segment σ_i. In our study, we selected the number of extracted topics to be between $k = 5$ and $k = 10$.

In a third and final step, for each segment σ_i the system combines the selected topics τ_i with a 1st-person view picture from participant's wearable camera captured anytime between the current segment's timestamp and that of the next segment (i.e., between t_i and t_{i+1}). The final picture selection is based on sharpness, number of faces captured (more is better), and the picture's *uniqueness* within the segments' time frame, as ascertained by a simple OpenCV[4]-based script we wrote. Each segment, and its set of topics, are packaged together with the corresponding picture into a single slide (see Figure 7.2 for an example of such slides). The system then assembles two customized slide decks (one for each participant, on average comprised of 20 slides) featuring the same set of topics but using pictures from each participant's perspective.

7.5 Study

In our study, we aimed for answering two research questions:

(i) How *much* can we improve participants' ability to recall a past meeting using automatically generated memory cues, based on semantic and contextual information captured during the meeting?

(ii) How does participants' recall of a past meeting influence their perceived effectiveness of a subsequent meeting?

For answering these questions, we used the previously described memory augmentation system that enabled us to present participants who meet regularly with a set of memory cues (in the form of a slide deck) to help them remember a past meeting. All participants had a string of at least 5 subsequent weekly

[4]`https://opencv.org`

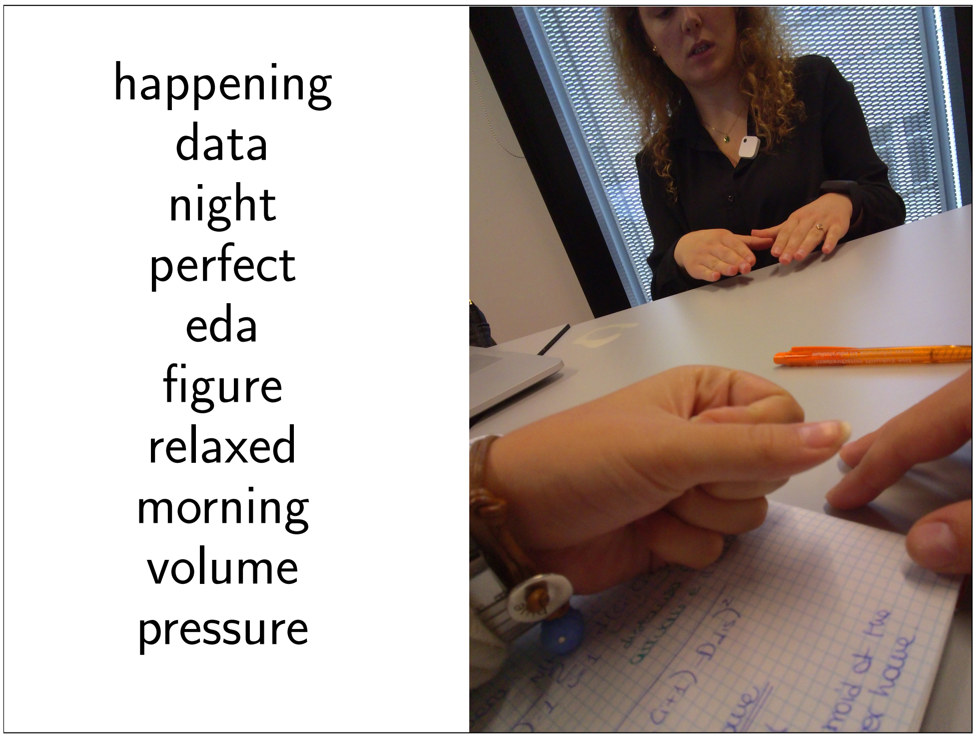


Figure 7.2. Example of a memory cue slide. In the experimental condition, participants reviewed about 20 such slides in order to improve their recall about the previous meeting. Slide's left: A set of 10 "topics" that were discussed during a particular "segment" of the previous meeting. Slide's right: A picture captured with the participant's Narrative Clip during the corresponding segment. Participants reviewed the same topics but saw always "their pictures", i.e., those captured from their point of view.

meetings, with each week being randomly assigned to one of the two following conditions:

- **A – Free Recall** (control condition): Right before an upcoming meeting, participants were individually interviewed about their recollection of their previous meeting. In particular, participants were separately asked to report whatever details they could remember from their previous meeting, while being voice recorded. In this condition, participants relied solely on their memory (no memory cues) for recalling the prior meeting.
- **B – Cued Recall** (experimental condition): Similar to condition A (free recall), participants were individually interviewed about what they could remember from their previous meeting. However, before the interview took

place, participants were asked to browse through a slide deck generated by our system, featuring both pictures and keywords from their previous meeting (see example in Figure 7.2). Pictures were taken from images of a wearable camera (i.e., Narrative Clip) that participants were asked to wear in all their meetings during the study. Keywords were automatically extracted from recorded audio transcripts of those meetings.

We used a within-subjects repeated design, performing each condition twice per participant (e.g., A-B-A-B or A-A-B-B) in a counterbalanced order to improve the reliability of our findings, with participants who met in fixed groups of two each time. For a total of 5 weeks of study with each group, we recorded 4 meetings and performed 8 pre-meeting interviews with our participants (no pre-meeting in week 1, no meeting after the interview in week 5, separate pre-meeting with each participant). Participants in the same group were always undergoing the same condition at each meeting.

We also formed the following hypotheses:

(H1) We expect when participants are exposed to our memory intervention (B – Cued Recall) will exhibit significantly higher memory recall of their previous meeting as opposed to when relying on their own memory (A – Free Recall). We attribute this to the combinatory power of utilizing hybrid memory cues that function synergistically for triggering both episodic memory (with pictures as memory cues), and semantic memory (with topics discussed as memory cues) [35, 48].

(H2) We also assume that the hypothesized memory augmentation effect achieved prior to a work meeting in condition B (Cued Recall), will be perceived as beneficial by our participants, increasing their perceived meeting effectiveness, as well as their reported valence and arousal at the end of the meeting.

7.5.1 Participants

We recruited a total of 12 participants (5 female) with an average age of 32.92 years (SD = 11.49 years) that met on an approximately weekly basis for a period of 5 weeks. Participants were recruited from our University and were either Professors, Ph.D., M.Sc., or B.Sc. students. Participants met in groups of two, in a goal driven (e.g., project/work progress) supervisor–supervisee fashion. We purposefully aimed for participants who meet in pairs to ensure that both sides are constantly engaged during the meeting. All groups discussed ongoing projects — ranging from project definition, to current progress and issues, to next steps. As two of our 12 participants participated in two groups (each being a group leader supervising two summer projects with a different student each), we had a total of 7 two-person groups. One group was tested for an additional week over condition B (i.e., 6 weeks in total). Two meetings had "guests", i.e., non-participants that would join for a single meeting due to project needs. Both participants and guests (for a single meeting only) signed informed consent forms acknowledging both the audio-visual recording of their meetings, and that audio recordings would be transcribed by a professional transcription service. All meetings were in English (which is the working language at our University), even though none of the participants was a native speaker.

7.5.2 Methodology

Each group of participants was tested four times in total for their ability to recall their previous meeting, twice for each condition over a period of 5 weeks. The first meeting for each group was always a *data collection*-only meeting (no prior meeting to recall yet), while the last meeting for each group was always a *recall interview*-only meeting (while participants had their regular meeting afterwards, we did not record this anymore). Participants belonging to the same group were always undergoing the same condition; conditions were counterbalanced across groups.

Data Capture

Just before a meeting would take place, we equipped both participants with *Narrative Clips* and *E4s*, making sure that the *Clips* were facing forward and were not obstructed by clothing, and that the *E4s* were activated and in contact with participants' wrists. We then set up the *HERO4* camera and tested that its field of view included both participants, while reminding participants that they could pause recording at any time using the *HERO4* remote. We then exited the room and participants were free to start their meeting. After they were done (typically within 30–60 minutes), participants would notify us to collect the equipment. At this time, we asked participants to additionally complete a short paper questionnaire each, inquiring into their meeting experience. In particular, participants used a printed version of the affect grid [191] to self-report on their valence and arousal levels during the meeting as well as, their perceived meeting effectiveness and recall session usefulness, in a scale from 1 ("not at all") to 11 ("very much").

Recalling the Last Meeting

Right before the next meeting, we individually interviewed both participants about what they could recall from their previous meeting. In condition B (cued recall), participants received 5 minutes time to review the custom slide deck we prepared, while in condition A (free recall) we simply asked them to think back about their past meeting (for a maximum of 5 minutes). In both conditions, participants were not allowed to take any notes during this 5-minute review period. However, we specifically instructed participants that they should follow their *regular note-taking routine* for all their meetings (some participants did so using paper notebooks or computers), as well as their *regular meeting preparation routine* prior to meeting with us (i.e., right before their next meeting). As participants were (potentially) following their routines for both of our conditions, our within-subjects designed ensured that we would still capture the relative improvement of our system for each participant. However, we did not explicitly verify if and/or how often they had consulted any such notes prior to our recall interview. After the 5 minutes were up, or the participants signalled that they were ready, we started a *recall interview*.

Recall Interview

The recall interview is intended to measure how accurately participants can remember their previous meeting after about one week, either with (condition B) or without (condition A) the support of our memory intervention. For this, we employed a specific interview technique known as the *cognitive interview*. The cognitive interview is an interviewing technique used by crime investigators for interviewing eyewitnesses [78]. In a cognitive interview, the interviewer constantly prompts the interviewee to recall more and more details (e.g., "*And then what happened?*") while taking notes. Once the interviewee is not able to recall any more details, the interviewer will go through the notes and read out different parts to extract additional details (i.e., "*Earlier you said... Can you tell me a bit more about that?*"). In our setting, this process not only served to improve the level of detail recalled by the participant, but also helped minimize misinterpretations. We recorded the recall interviews prior to the subsequent meeting of our participants in a quiet office, hence we could rely on using only a voice recorder to allow our 3rd party transcription service generate a high-quality transcription[5] of a participant's recall.

[5] Note that these transcripts were *not* used to create memory cues, these were only used to measure participant recall.

Coding and Rating Recall

In order to assess each participant's recall of a meeting, we followed a two-stage process: First, we manually assessed what topics were discussed during each meeting (coding step). Then, we manually matched these topics against what each participant remembered from the meeting (rating step).

Coding processed as follows: for each meeting transcript, one researcher would extract all discussed topics using a 3-level hierarchy, i.e., main topic, subtopic, and sub-subtopic. We linked all identified topics to its corresponding quotation using Atlas.ti[6], ultimately exporting all topics and (sub-)subtopics in a meeting into a *codebook*. In order to ensure that we identified the correct codes in this step, a second researcher would then take a random sample of meetings (20 %, i.e., 6 out of 30) and use this codebook to independently tag its contents. Typically, a 10 % subset size is enough for assessing inter-coder reliability [38]. We compared the two different topic connection schemes from both researchers, using a variant of Krippendorf's Alpha inter-coder reliability test, called multi-valued coding [136]. Normally, an alpha value above 80 % indicates good inter-coder reliability [241]. The test revealed an average inter-coder reliability between the two researchers at the level of 80.553 % (SD = 10.037 %).

After the codebook for each meeting had been finalized, the two researchers would independently *rate* how accurately participants were able to recall each past meeting, based on their respective recall interviews. In this step, the researchers linked a meeting's codebook topics with the corresponding text segments in two transcribed recall interviews about that meeting. Thus, each researcher produced two ratings per meeting, one for each participant. We then averaged these two scores for each participant, resulting in an overall recall rating per participant about a given meeting. For testing the inter-rater reliability, we simply calculated the percentage of overlapping topic matches between recall interviews and meeting transcripts for the two researchers. The analysis displayed an average inter-rater reliability of 96.436 % (SD = 2.967 %).

7.6 Results

During the study, we collected a total of 29 meeting transcripts (and corresponding 58 recall interview transcripts) from a total of 7 groups of participants who, for 5 weeks, met roughly every 7 days. For week 5, we simply interviewed our

[6] `http://atlasti.com`

participants about their previous meeting and we concluded the study without recording their upcoming meeting. One group was tested for an additional week over condition B, thus producing a total of 29 meeting transcripts instead of the expected 28. Meetings lasted on average 43.316 minutes (SD = 19.683 minutes) while recall interviews lasted on average 8.933 minutes (SD = 4.183 minutes). The *5 minutes maximum* recollection sessions lasted on average 58 seconds (SD = 61.971 seconds) for condition A, and 3.35 minutes (SD = 1.505 minutes) for condition B. Meeting transcripts had on average 5379.71 words (SD = 2589.259 words), while interview transcripts featured on average 1077.98 words (SD = 546.183 words). We extracted an average of 41.41 topics (SD = 21.6, subtopics and sub-subtopics included) discussed for each meeting.

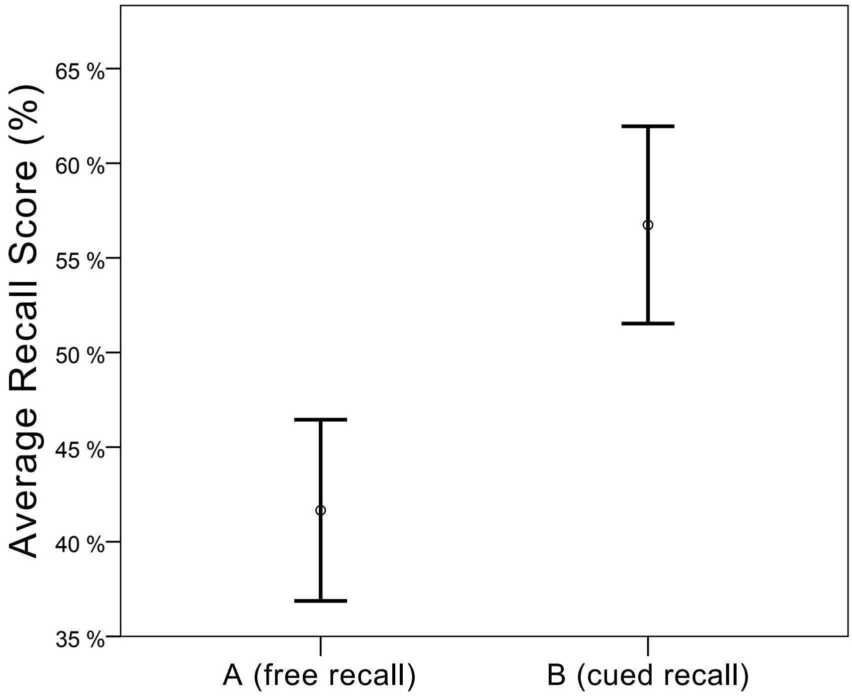


Figure 7.3. Average recall of the previous meeting during condition A (free recall) and condition B (cued recall). Using our memory intervention (i.e., condition B), participants were on average able to recall 15 % more than when relying solely on their own memory (i.e., condition A).

Our study design produced 4 recall scores per participant per meeting — two for condition A and two for condition B. Recall scores within the same condition differed on average by 10.744 % (SD = 6.492 %) for condition A (free recall), and by 12.989 % (SD = 13.666 %) for condition B (cued recall). A paired-samples t-test between the score differences for condition A ($|A1–A2|$)

and condition B (|B1–B2|) allowed us to assess the consistency of recall scores within the same condition, and displayed no significant difference ($t(13) = -.64$, $p = .533$). This indicates that participants were generally consistent in what they could recall about their previous meeting over the two times they were tested for both conditions. To test if our intervention influenced our participants' ability to recall their previous meeting, we conducted a paired-samples t-test to compare between the overall recall scores achieved in condition A (free recall) and in condition B (cued recall). This showed a significant difference in the overall recall scores for the two conditions ($t(13) = -5.719$, $p < .001$), indicating that participants' overall recall scores in condition B ($M = 56.74$ %, SD = 9.025 %) were significantly higher than participants' overall recall scores in condition A ($M = 41.661$ %, SD = 8.292 %). On average, participants were hence able to recall 15 % ($M = 15.079$ %, SD = 9.865 %) more details (i.e., topics discussed) about their previous meeting when using our memory intervention (see Figure 7.3), as previously hypothesized (H1). This finding answers our first research question (i) by quantifying a baseline by which we can improve participants' recall of a prior work meeting using automatically generated memory cues.

Based on the self-reported valence and arousal levels during a meeting, as well as the perceived meeting effectiveness and recall session usefulness (all indicated by participants in our post-meeting paper questionnaire), we also inquired into the perceived effects of memory augmentation on the actual meeting. However, a series of Wilcoxon signed-rank tests revealed no significant differences in reported valence ($Z = -.856$, $p = .392$) and arousal ($Z = -.09$, $p = .928$) levels, as well as in perceived meeting effectiveness ($Z = -.67$, $p = .503$) and recall session usefulness ($Z = -.276$, $p = .783$), between condition A and B. These findings indicate that, even though our memory intervention did significantly improve participants' ability to recall their previous meeting, it did not significantly affect their self-reported emotional state or how effective they felt their subsequent meeting was (ii), nor did they feel that the recall session using slides (condition B) was more useful than the free recall condition (A) hence, contradicting our previous hypothesis (H2).

7.7 Discussion

Reviewing a slide deck with past meeting topics displayed, will *obviously* help to increase one's recall of that meeting. The main goal of our experiment was thus not to show that such an increase exists, but to more concretely ascertain the *actual gain* that the combination of episodic and semantic memory cues can attain. Several follow-up questions arise. Firstly, how much of this increase can be attributed to the components of our memory cues? Our slide deck used both *semantic* and *episodic* memory cues (i.e., topics and pictures). While we did not examine how each cue worked in isolation, informal feedback from our participants confirmed that they mostly relied on the semantic cues (i.e., the extracted topic keywords) when refreshing their memory of a past meeting. We attribute this to the sedentary setting of a work meeting, in which participants were mostly seated around a table and not moving, hence captured images across the entire experience were not very discernible. On-site meetings (e.g., inspecting a building site) or more active meetings (workshop style) might provide a richer visual experience, hence improving the power of episodic memory cues (pictures). We also did not test how simple note-taking (on paper, on a computer) would compare, nor how simply having a full transcript might already help. While either of these two interventions might produce equal recall improvements, both would be more laborious: note taking requires extra effort during the meeting, while reading a full transcript (typically around 10 pages) would certainly take more time than reviewing a few slides. Clearly, being able to concretely compare the recall gain of our memory augmentation system against those alternatives would be very valuable — it is something we plan to do in the future. We are also planning to analyze the physiological data we collected during the meetings, in order to better understand how individual arousal and alertness can help predict those moments in a meeting that have a high potential for acting as memory cues.

We were surprised to find no significant differences in the subsequent meeting's effectiveness (as indicated by the reported valence and arousal levels, subjective meeting effectiveness, and perceived usefulness of recall session) between the two conditions. We expected that the increased recall in our experimental condition would also result in higher meeting productivity and effectiveness (H2). It could very well be that the subjective assessment of these measures is much more influenced by the actual topics discussed in a meeting, than by how much one remembered from a past meeting.

Our participants rarely used up the maximum of 5 minutes that we had set for our recall sessions. The 15 % average memory improvement score in condition B was achieved by participants reviewing slides for an average of only 3.35 minutes. This suggests that review effectiveness, in particular when comparing with a hitherto untested memory aid, such as a full transcript or the output of an automated text summarizer, may be an important aspect to explore. Another important parameter to further investigate is also recall persistence, i.e., how long people will remember meeting contents after having seen the summary. A future experiment would need to measure both the effects of *time after review* and *number of reviews* on memory recall (i.e., spaced repetition [9] — see Chapter 3). Finally, while we relied on human transcription services to provide the core input to our memory augmentation system, advancements in natural language processing will eventually enable the fully automated, end-to-end (i.e., from capture to presentation) provision of effective memory cues.

7.8 Summary

Our five-week study with 12 participants allowed us to quantify the memory improvement of which our automated memory augmentation system is capable: 15 % on average. Interestingly, this memory improvement was achieved with participants spending an average of only 3.3 minutes reviewing the provided memory cues. These figures can act as a first baseline against which to evaluate alternative methods of memory augmentation and future pervasive memory augmentation systems. While informal feedback showed that participants did not seem to find the pictures in our slide deck all that helpful, we nevertheless believe that the power of our system lies in the combination of both episodic and semantic memory cues. One interesting option may thus be to apply our system in situations where such a combination of memory cues may be more helpful, for example in learning environments such as schools or universities, where (*semantic*) knowledge acquisition is thoroughly intertwined with (*episodic*) experiences.

All in all, this field study corroborated prior evidence on the potential of cue-based memory augmentation for producing tangible memory benefit in the settings of everyday life. Notably, our findings highlight the potential of hybrid memory cues to synergistically trigger both semantic and episodic memory recall (i.e., RQ3) from which future pervasive memory augmentation systems can benefit. In the next chapter, we explore the aptitude of physiological responses, captured with commercial wearable devices, to inform memory cue selection when it comes to recalling a past virtual reality experience, and hence potentially further improve memory cue quality for recalling past physical experiences too.

Chapter 8

Physiological Responses in VR Experience Recall

In this chapter, we investigate the potential of physiological responses to indicate significant moments to capture and replay later, in the form of **memory cues**. We present early results from two studies, and a prototype that attempts to incorporate physiological responses in automatic picture capture, as a sort of "**memory biomarkers**", indicators of strong (or weak) memory formation. These so-called "memory biomarkers" could be utilized for guiding the capture, selection/generation, and/or the presentation of memory cues, and hence drastically improving overall memory cue quality (i.e., RQ4). In the following sections, we present the motivation behind using physiological responses and picture capture for augmenting memory recall. We describe two completed studies, one in a museum, and one in an aquarium, where we explore the role of physiological responses and VR scene characteristics in the recall of past VR experiences. We introduce the concept of physiologically–driven memory cue selection and we illustrate our "PulseCam" mobile and wearable application prototype, which captures pictures based on one's heart rate, a feature that could be incorporated in future pervasive memory augmentation systems.

8.1 Author's Contribution

The author of this thesis had a leading role in the studies reported in this chapter. In particular, his contribution includes the conceptualization, study design, data analyses, and writing. The supporting publications [164, 167] reported in this chapter aim primarily at the field of tourism, and hence are not entirely related to the field of human memory augmentation. In addition, a significant portion of the analyses and results, reported in this chapter, have not been published yet. Nevertheless, the contribution of the co-authors in the supporting publications was integral, since they conducted the user study and collected the data. For more information on the supporting publications, see [164, 167].

8.2 Introduction

Various studies in Neuroscience and Psychology have shown a correlation between arousal during an attended event and memory recall of details related to that specific event. Viewing arousing stories, for example, can result in experiencing a greater emotional reaction to the story than when viewing neutral stories, which subsequently results in enhanced memory for the story [96]. The intensity of an emotion experienced by a person (i.e., arousal) was found to account for significantly more variance in autobiographical memory characteristics than did valence or age of the memory [138, 221, 240]. Emotional memories result in better recall of perceptual, sensory, and semantic elements of an event in comparison with neutral memories [203]. Hence, one would expect that by monitoring one's physiological responses it is possible to approximate the moment at which strong memories are formed during an experience. However, physiological responses are notorious for fluctuating due to a plethora of reasons (i.e., physical activity, induced stress, and other external stimuli), essentially rendering impossible the ascription of physiological activity to memory formation. Nevertheless, Virtual Reality (VR) may provide the ideal opportunity for testing if we can detect memory formation via physiological responses monitoring and what visual/audible stimuli (i.e., VR content) could cause it. VR experiences are generally exhibited using head–mounted displays (HMD) that allow one to fully immerse oneself into the experience, and in combination with earplugs, may eliminate all external stimuli (e.g., sounds).

8.3 Recalling Virtual Experiences

Recently, Virtual Reality (VR) has seen increasing adoption in the tourism domain and particularly museums. VR is used in museums to enrich the traditional exhibit viewing experience with simulating settings that cannot be recreated or simply visited. For example, a natural history museum, can offer a VR experience that approximates a setting during the Jurassic period. Hence, museum visitors may witness a far more vivid experience than just viewing stationary dinosaur fossils. Furthermore, museums are highly visited locations by people that are generally relaxed and in the mood to explore, learn, and experience something new. We thus decided to investigate the potential of physiological responses to regulate memory formation during a VR experience in the context of museums. In particular, we strive for answering the following research questions:

i. *Can we predict what participants will remember of a VR experience based on their physiological responses during the experience?*
ii. *Which are the elements that make a VR experience particularly memorable?*

For answering our research questions, we recruited participants to go through already provided VR experiences in two highly visited exhibitions: A) the **museum** of Saint Gotthard in Switzerland, and B) the **aquarium** of Genova in Italy. The first VR experience was viewed by a total of 23 participants, and it was animating the construction and opening of the new Gotthard tunnel under the Alps. The second VR experience was viewed by a total of 46 participants, and was about a virtual dive into the deep sea. In the first case, participants were recruited prior to the experiment, whereas in the second, participants were recruited in situ. In both cases, participants were asked to pre-fill a questionnaire obtaining their consent, then sat to a chair, wore the E4 wristband and the VR Head Mounted Display (HMD). Next, participants simply viewed the corresponding VR experience while seated and while their physiological responses were measured. Upon finishing, participants completed a post-questionnaire inquiring, among others, into their favourite scenes and moments during the VR experience they just encountered. A month later, and in both cases, participants were contacted via e-mail or telephone, and were asked to recall about the corresponding VR experience they had viewed, and report any detail they could remember.

We also formed the following hypotheses:

(H1) We expect participants' recorded physiological responses will exhibit significant variations as a result of being fully immersed into the VR experience. A VR experience, similarly to a physical experience, should be able to induce emotions to our participants that will in turn be exhibited as fluctuations in their physiological responses [198]. Subsequently, exhibited fluctuations in participants' physiological responses can be captured via physiological monitoring equipment, such as the one we use in this work.

(H2) We expect that participants recorded physiological responses during a VR experience will display a relationship to which elements of the VR experience participants can recall later. Prior work has shown that physiological responses such as EDA (Electro-Dermal Activity), has been successfully utilized for selecting pictures that better support episodic memory recall [200]. Hence, we expect we will be able to find a similar relationship between participants' VR recollections and additional exhibited physiological responses, beyond just EDA.

(H3) We also assume that certain aspects of a VR Scene, such as unusual perspective, sudden changes, animations, and sounds, will have an effect on which parts of the VR experience participants can recall later. Similarly to film viewing, we expect these VR aspects would cause "intrusive memory formation" [113], that in turn would lead participants to better recall certain VR Scenes.

So far, we have synchronized the physiological data for the total of 69 participants we recruited. By dividing each VR experience in Scene segments, we were able to detect systematic variations in participants' average Heart Rate (HR) and Heart Rate Variation (HRV) levels, in both VR experiences. We have also extracted generic features for each Scene (i.e., perspective, animations, movement, etc.) for investigating their role on the recorded physiological responses. In the case of the museum VR experience, we have collected participants' recollections in an attempt to examine if there is any association with the fluctuations of their recorded physiological responses. Early findings and results show that HR, as measured with E4, displayed significant variation during the course of the VR experience on average, and have already been published in [153, 154].

8.3.1 VR in the Museum

The VR experience investigated in this study was about the tourist attractions of Ticino, the southernmost Canton of Switzerland. The VR experience was developed by Ticino Tourism[1], the related Swiss regional Destination Marketing Organization (DMO), with the intention to use it as a marketing material for promoting Ticino as a destination, in the occasion of the Gotthard tunnel opening in June 2016. The VR experience[2] lasts for 5:45 min and is comprised of 11 scenes of unequal duration, featuring a virtual journey through the Gotthard tunnel to typical Ticino sights and landscapes. By the time the museum VR experience was studied, it had already been introduced to the public in several events organized by the DMO, and it had been viewed for 12,857 times (June–August 2016). In the following sections, we briefly present our results and discuss our

[1] http://ticino.ch
[2] http://www.ticino.ch/en/campaigns/alptransit/vr.html

early findings. For a more in-depth view on the study and the literature review, see [153, 154].

Study

We recruited our participants through a post on the official Facebook page of the DMO, inviting people from central Switzerland to test the new Ticino VR experience. We deliberately opted out of recruiting locals for limiting the effect of familiarity on our findings. In total, we recruited 23 people (4 females and 19 males) for participating in the experiment. The experiment took place on June 8, 2016 in Zurich, Switzerland. Each participant was invited at a pre-arranged time in an office hosting the experiment and two Oculus Rift HMDs. After having signed a consent form, each participant was equipped with the Oculus Rift HMD and the Empatica E4 wristband for recording participant's physiological responses, such as Heart Rate (HR) and Electrodermal Activity (EDA), in-sync with the VR experience displayed by the HMD. In addition, our participants were administered with three questionnaires: one before the experience, one immediately after, and one approximately a month after the VR experience, inquiring, among others, into which parts of the VR experience our participants could recall.

Data Collection and Analysis

We managed to collect a dataset of considerable volume about our participants' real–time physiological responses when encountering the VR experience under study. In particular, the raw dataset is comprised of Blood Volume Pulse (BVP) recorded at 64 Hz, Inter-beat Interval (IBI) when occurred, Electrodermal Activity (EDA) at 4 Hz, Skin Temperature at 4 Hz, and 3D Acceleration sensor readings at 32 Hz. Synchronizing all these different types of data streams is non-trivial, particularly over large sampling periods and when it comes to syncing with an externally added stream, resource or intervention, which in this study is the VR experience per se. After all data has been synced, we perform data reduction techniques, removing outliers or data that do not correspond to meaningful experience segments. The next step is feature extraction, where we calculate additional meaningful variables such as Heart Rate (HR) and peaks (e.g., HR peaks, EDA peaks, temperature peaks, etc.). Participants' transcribed recollections were thematically processed for extracting overall recall scores and episodic richness scores for each VR Scene [146]. In addition, we performed a content analysis of the elements displayed in each scene of the VR experience, eliciting and classifying the components of the displayed VR media characteristics. Guided by literature review on VR in Tourism, available in [153, 154], we identified and coded the following four aspects for each VR Scene:

- **Format of the scene.** The scene might be displayed as a 360° photo or as a 3D video reconstruction.
- **Horizon perspective.** The view of the scene can be from a regular perspective (eye–level perspective), or from an unusual perspective. As an unusual perspective we consider a tilting of the horizon lines upward or downward.
- **Animated elements in the scene.** The scene contains only static objects or includes also animated elements moving within the scene.
- **Sound effects.** the scene may have no sound at all, a music background, or a combination of sound alterations (e.g., sound of a train passing by).

Results

Most of the participants had already visited the destination (20 out of 23), whereas almost half of them (11 out of 23) had a prior experience with a VR HMD and a VR experience. Participants enjoyed in general the VR experience, reporting the following elicited emotions: happiness, excitement, being impressed and surprised. The experience has been primarily classified as informative, entertaining, and playful. The questionnaire administered 1 month after the experience had taken place, collected 18 answers out of 23 participants addressed in total. Participants recalled the experience as very positive, informative, and entertaining. In this section, we combine participants' physiological responses, recorded during the VR experience, with participants' questionnaire responses and recollections for investigating the effect of VR Scene and their characteristics on participants' arousal and recall, and the relationship between them.

Scene Effect on Heart Rate

Participants' average heart rate (HR) levels exhibited notable fluctuations throughout the museum VR experience (see Figure 8.1). In particular, HR levels increased rapidly on average during Scenes 1, 2, and 3, gradually decreasing during Scenes 4, 5, 6, and 7, before increasing again during Scenes 8 and 9. For systematically investigating the effect of Scene on participants' HR levels, we performed a mixed–design analysis of variance (i.e., mixed repeated measures ANOVA), with participants' HR as dependent variable, VR Scene as an independent variable, and time (in seconds) as a repeated measures factor. Mixed–design

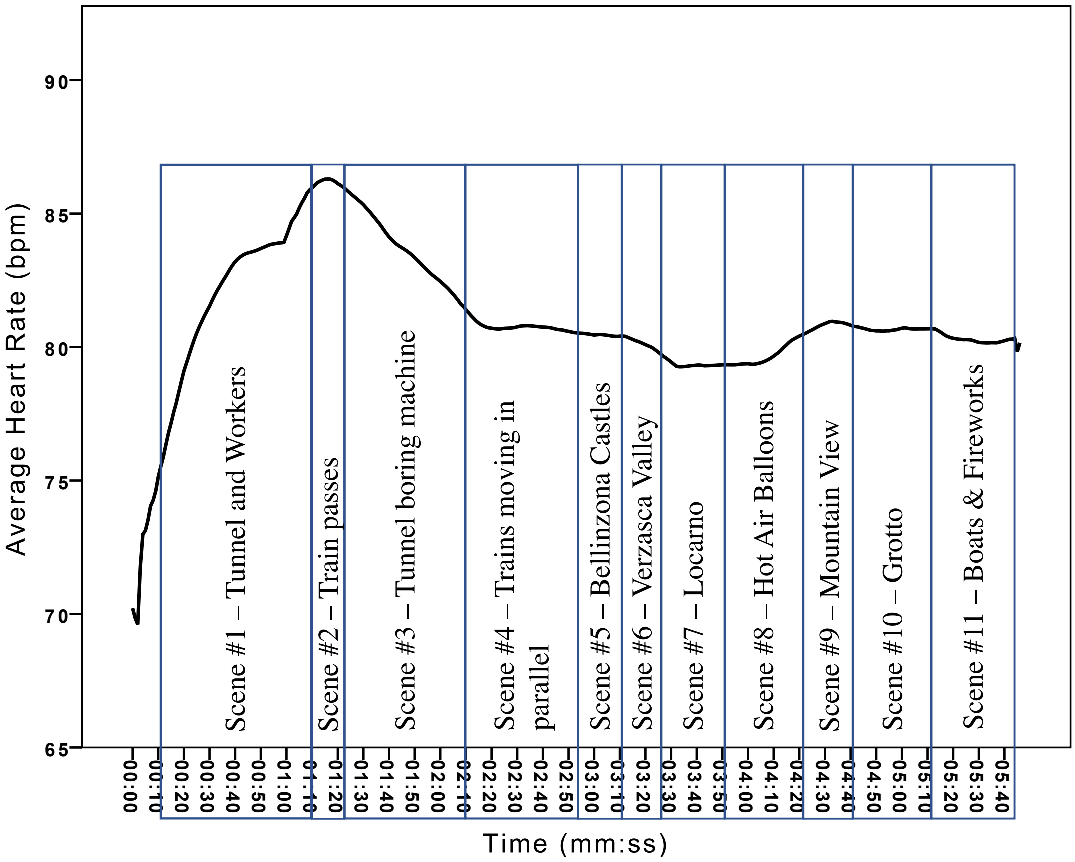


Figure 8.1. Average Heart Rate (HR) variation during the museum Virtual Reality experience. Participants during Scenes 1, 2, and 3 displayed significantly higher HR levels than in all other Scenes.

ANOVA bears a significant advantage over classic repeated measures ANOVA procedure, in that it accounts for unequal number of cases across the different levels of the independent variable. Thus, we were able to compare physiological responses corresponding to VR Scenes of unequal duration (e.g., 59 seconds of HR measurements for Scene 1 vs. 12 Seconds of HR measurements for Scene 2), without performing random case selection, multiple imputation [163], or even pairwise exclusion, that would have dramatically decreased our sample size. The analysis displayed a significant main effect for VR Scene on participants' exhibited average HR levels ($F(10, 7964) = 23.122$, $p < .001$). Post-hoc pairwise comparisons using the Bonferroni correction showed that participants during Scene 2 ($M = 86.162$, $SE = 2.766$, $p < .001$ in overall, and $p < .05$ against Scene 3) displayed systematically and significantly the highest estimated average HR, as opposed to all other VR Scenes. The similar trend is observed during Scene 3 ($M = 83.625$, $SE = 2.728$, $p < .001$ in overall, and $p < .05$ against Scene 1), where estimated average HR displayed significantly higher levels in comparison to all other Scenes, apart for Scene 2. Last but not least, participants' estimated average HR during Scene 1 ($M = 82.18$, $SE = 2.724$, $p < .05$), exhibited significantly higher levels, as compared to all other Scenes, apart from Scene 2, 3, and 9. For an overview of the average estimated HR for all VR Scenes see Table 8.1. These findings corroborate our hypothesis (H1) in that the VR experience will have an impact on participants' physiological responses.

VR Scene Nr.	Estimated Mean (HR)	Std. Error (HR)	Total Recalls (right after)	Total Recalls (1 month after)
1	**82.180**	**2.724**	14	4
2	**86.162**	**2.766**	10	7
3	**83.625**	**2.728**	10	3
4	80.753	2.729	10	7
5	80.456	2.752	5	7
6	80.083	2.756	3	5
7	79.332	2.742	5	8
8	79.701	2.735	7	12
9	80.823	2.750	0	9
10	80.667	2.735	15	11
11	80.265	2.734	4	10

Table 8.1. Overview of estimated marginal average Heart Rate (HR) in bpm, standard errors (SE) per Scene, and total number of times a scene was recalled in a follow up interview, right after and a month after the VR experience. The Scenes with the highest observed HR appear in bold font.

At this point, we need to note that we expect, at some extent, a "*carry–over*" effect on currently measured participants' HR levels induced by previous temporally adjacent scenes. In other words, we expect that the HR levels exhibited and measured continuously in a previous Scene may influence the HR levels measured in the next Scene. However, for preventing such a phenomenon from manifesting, we would have to pause the VR experience every time a VR Scene had ended, let participants calm down for some period (e.g., 5 min), and continue with the next Scene. Nevertheless, this was not considered a viable option, both from the perspective of time, and that of quality of experience. In fact, a heavily segmented VR experience would have been perceived as disruptive and unpleasant by our participants. Therefore, we expect for example, that participants' average HR levels during Scene 3 may have influenced participants' average HR levels measured during Scene 4.

Heart Rate Peaks and Memory Recall

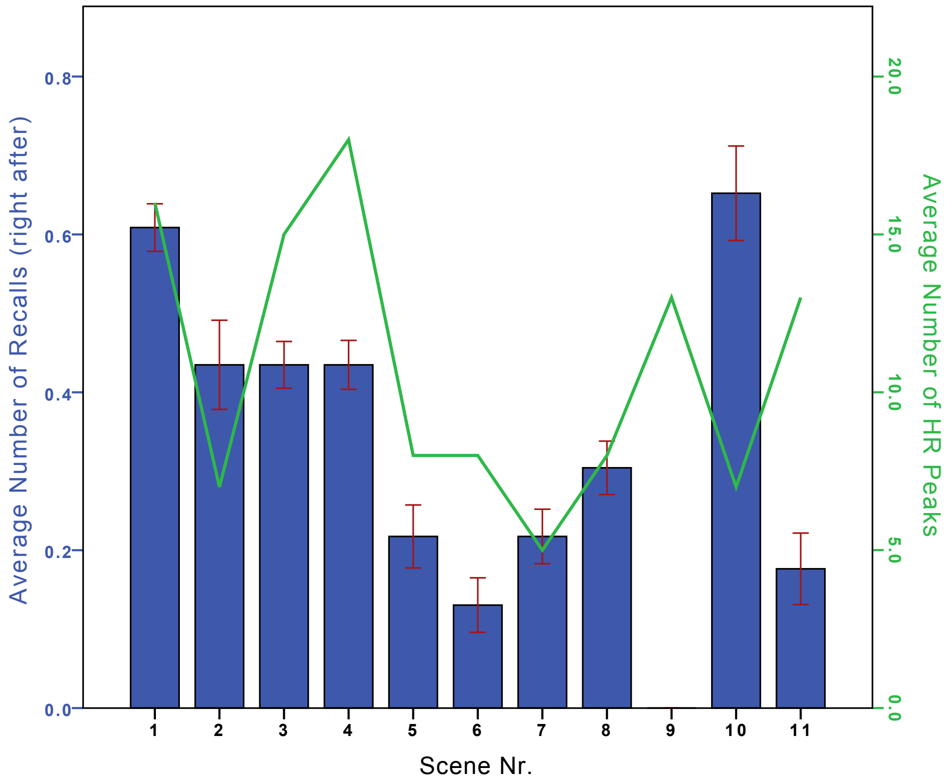


Figure 8.2. Average number of times a VR Scene was recalled, right after the VR experience, in relation to average number of heart rate (HR) peaks. A significant positive correlation was found.

So far, we were able to calculate features such as EDA and HR peaks as additional indicators and regulators of emotion and arousal [5]. As far as HR peaks are concerned, we were able to detect significant correlations with the number of times a VR Scene was recalled, right after and a month after the VR experience was encountered. In particular, a significant positive Spearman rank-order correlation was found between average number of recalls per Scene, right after the end of the VR experience, and the average number of HR peaks per Scene ($r_s = .15$, $N = 7717$, $p < .001$) (see Figure 8.2). Interestingly, the opposite trend

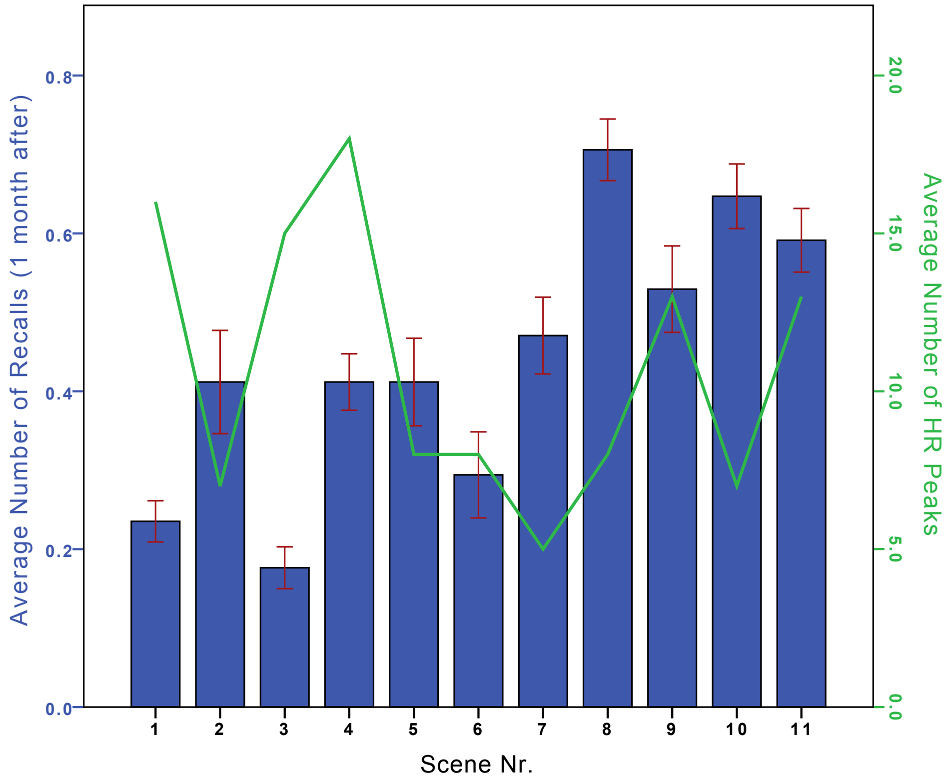


Figure 8.3. Average number of times a VR Scene was recalled, one month after the VR experience, in relation to average number of heart rate (HR) peaks. A significant negative correlation was found.

is observed when examining the Spearman rank-order correlation between average recall count per Scene, a month after the VR experience, with the average number of HR Peaks per Scene ($r_s = -.197$, $N = 5709$, $p < .001$) (see Figure 8.3). These results indicate a possible interplay between participants' exhibited physiological responses (in terms of HR peaks), and their ability to recall an experience at subsequent stages. These findings provide ground for our initial hypothesis (H2) in that participant's physiological responses recorded during the VR experience will display a relationship with which parts of the VR experience participants can recall later. However, further analyses (e.g., EDA peaks) are needed for reliably supporting this claim.

VR Scene Characteristics and Recall

When investigating the media characteristics of the most recalled scenes, as reported by the participants right after and a month after the museum VR expe-

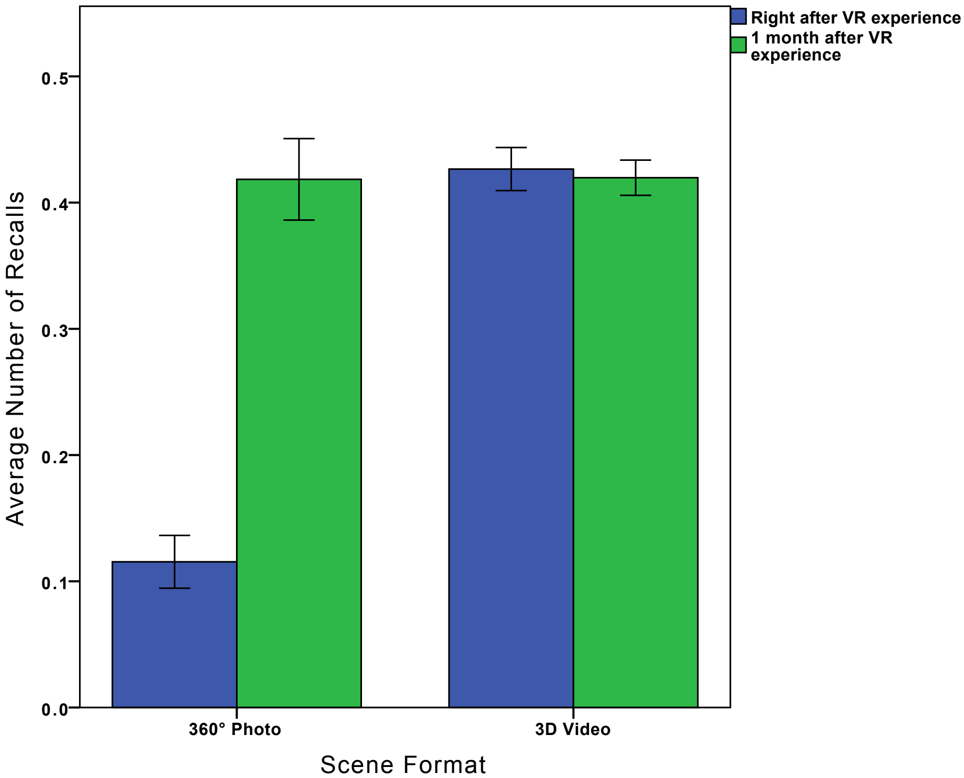


Figure 8.4. Average number of VR Scene recalls, right after the VR experience and a month after, in relation to Scene format. A significant difference was found in the number of recalls right after the VR experience.

rience, it is evident that the top recalled scenes were the ones characterised by an unusual horizon perspective (namely Scene 8 and 10). In fact, in both cases, the view of the Scene was introduced from an unusual perspective: Scene 8 simulates a flight on a hot air balloon, while Scene 10 presents a landscape view from the balcony of a restaurant located on top of a mountain. In both Scenes 8 and 10, animated elements are displayed and a particular sound (e.g., wind blows when on the balloon) denotes the "mood" of the scene. This suggests that media characteristics may have had an effect on participants' ability to recall a Scene. For investigating this hypothesis more systematically, we conducted two one–way analyses of variance (ANOVA) with number of times a VR Scene was recalled right after and one month after the VR experience as dependent variables, and the Scene format (360° photo vs. 3D video). The analyses displayed a significant main effect for Scene format on how many times a Scene was recalled right after the experience ($F(1, 7726) = 341.327$, $p < .001$, $\eta_p^2 = .042$), but no significant main effect on how many times a VR Scene was recalled a month after the VR experience ($F(1, 5729) = .001$, $p = .972$, $\eta_p^2 = .0$). This indicates that our participants were significantly more able to recall a VR Scene featuring a 360° photo a month after the VR experience than they were right after the VR experience. Interestingly, the same effect was not found for VR Scenes that feature a 3D video. These results suggest that the VR Scene format plays a significant role in one's ability to recall a VR Scene, and in turn a VR experience, in later stages of recall, as previously hypothesized (H3). In fact, additional media characteristics we have elicited, such as horizon perspective, animated elements, and sound effects, may also play an important role on one's ability to recall a VR experience, but further analyses should be conducted.

8.3.2 VR in the Aquarium

The second VR experience we investigated was the "descent into the abyss", in a virtual submarine, available at the largest aquarium in the world: the Aquarium of Genova[3], in Italy. This immersive VR experience engrosses one in the aquatic environment of the deep sea, and through encounters with its strange creatures, introduces and educates one about the unearthly, deep ocean world. Same as with the museum VR experience, the aquarium VR experience was publicly available via a set of Oculus Rift HMDs, but hosted in a dedicated "Abyss Room"[4], designed for further facilitating immersion during the experience (e.g., special lighting conditions, quiet room, 360°–rotating seats, etc.). The aquarium VR experience lasts in total for 2:51 minutes and is comprised of 6 VR Scenes of unequal duration.

Study

We visited the Aquarium of Genova for two consecutive days (August 22 and 23, 2016), and recruited volunteers in situ from the large pool of the daily visitors, who in their large majority were tourists. In particular, we approached those visitors that were interested in the provided VR experience and asked them whether they would be willing to participate in a small study involving a pre- and a post-questionnaire as well as, the use of the Empatica E4 wristband while encountering the aquarium VR experience. In total, we were able to recruit 46 participants, after obtaining their consent about collecting and analysing their physiological

[3] https://www.acquariodigenova.it/en/
[4] https://www.youtube.com/watch?v=Pofg-jTOoEc

and questionnaire responses. All participants encountered the aquarium VR experience for the first time during our study and hence, we do not expect any familiarity effects. However, despite we also collected our participants' contact details (e.g., e-mails), we were not able to elicit a sufficient number of responses in a follow-up invitation over e-mail a month later (only 8 our for 46 responded), asking participants to send in their recollections about the aquarium VR experience. Nevertheless, we were able to unveil significant fluctuations in participants' physiological responses during the VR experience.

Data Collection and Analysis

Similarly to the museum VR experience, we have used Empatica E4 to collect our participants' physiological responses (i.e., BVP, IBI, EDA, Skin Temperature, and 3-axis Acceleration) during the aquarium VR experience. For estimating the effect of a VR Scene on the collected physiological responses (e.g., HR and HRV), we first group these data (i.e., dependent variable) by the levels of our independent variable VR Scene. Then, we proceed to the analyses with a mixed-design analysis of variance (i.e., mixed repeated measures ANOVA). There is a range of reasons we selected mixed repeated measures ANOVA for analysing the effect of Scene on HRV. First, a mixed-design approach allows the comparison of levels/-conditions with unequal samples and as such, we were able to compare Scenes of unequal duration (similarly to the museum VR experience). Second, our sample is comprised of physiological measurements collected in predefined intervals, where one cannot assume independence of observations or homogeneity of variance across our measurements (i.e., necessary ANOVA assumptions). Fortunately, specific types of repeated covariance such as "Ante-Dependence: First Order" is robust for heterogeneous variances and heterogeneous correlations thus, particularly suitable for our sample.

Results

In this section, we first report the additional features we calculated from the aquarium physiological responses dataset, and we proceed in determining the extent to which the physiological responses were affected by each separate VR Scene.

Scene Effect on Heart Rate Variability

Inter-Beat Interval (IBI) or "beat-to-beat" interval refers to the time interval between individual heart beats. IBI is the measure used for expressing Heart Rate

Variability (HRV). Due to the nature of IBI measure (i.e., heart beats occur arbitrary outside predefined time intervals), there is no sampling rate one can use for synchronizing IBI output with the rest of E4 physiological measurements (e.g., HR, Temperature, EDA, etc.). Thus, for synchronization purposes, we defined a new variable describing HRV produced from IBI in 1 Hz rate. For this, we defined a 1-sec time window during which we sum all IBI falling in this window. The lower the IBI the higher the HRV (i.e., $HRV = \frac{1}{IBI}$). We describe this below in a more formal fashion:

Let t_i be the timestamp in milliseconds at which a heartbeat was registered and $f(t_i) = \frac{t_i}{1000}$ the function that converts milliseconds to seconds. Also, let $T = \{t_0, ..., t_N\}$ be a set of timestamps where $\forall t_i, t_j \in T$ and where $f(t_i) = f(t_j)$.

$$\text{Then, } HRV_{f(t_i)} = \frac{1}{\sum_{i=0}^{N-1}(t_{i+1} - t_i)} \tag{8.1}$$

We then calculated the HRV for all our participants. We first pre-processed our HRV data removing outliers and empty values for the timestamps that no PPG signal was obtained. This limited our sample to a size of $N = 5050$ valid HRV measurements and we checked its distribution. Since HRV distribution was not deviating significantly from normal (Skewness = .852 and Kurtosis = −.51), we did not apply any normality transformation. Then, we grouped our participants' exhibited HRV by 6 distinct Virtual Reality (VR) Scenes of unequal duration (see Figure 8.5). A Levene's test of homogeneity of variance did not confirm the assumption of equal variances across all the levels of the independent variable Scene ($F(5, 5044) = 4.436$, $p < .001$). For estimating the effect of VR Scene on participants' recorded HRV, we performed a one-way mixed–design, repeated measures ANOVA, with participant' HRV as a dependent variable and scene number as an independent variable. However, the analysis displayed a non–significant main effect of Scene on participants' exhibited HRV ($F(5, 510.927) = .550$, $p = .738$). This finding contradicts our initial hypothesis (H1) in that the VR experience will have an impact on participants' physiological responses.

Scene Effect on Heart Rate

We first pre-processed our Heart Rate (HR) data removing outliers limiting our sample size to $N = 7495$ valid HR measurements. After checking its distribution and finding some variations from normal, we decided to normalize our HR data with a two–step approach for transforming continuous variables to normal [223]. A Kolmogorov–Smirnov normality test verified the assumption of

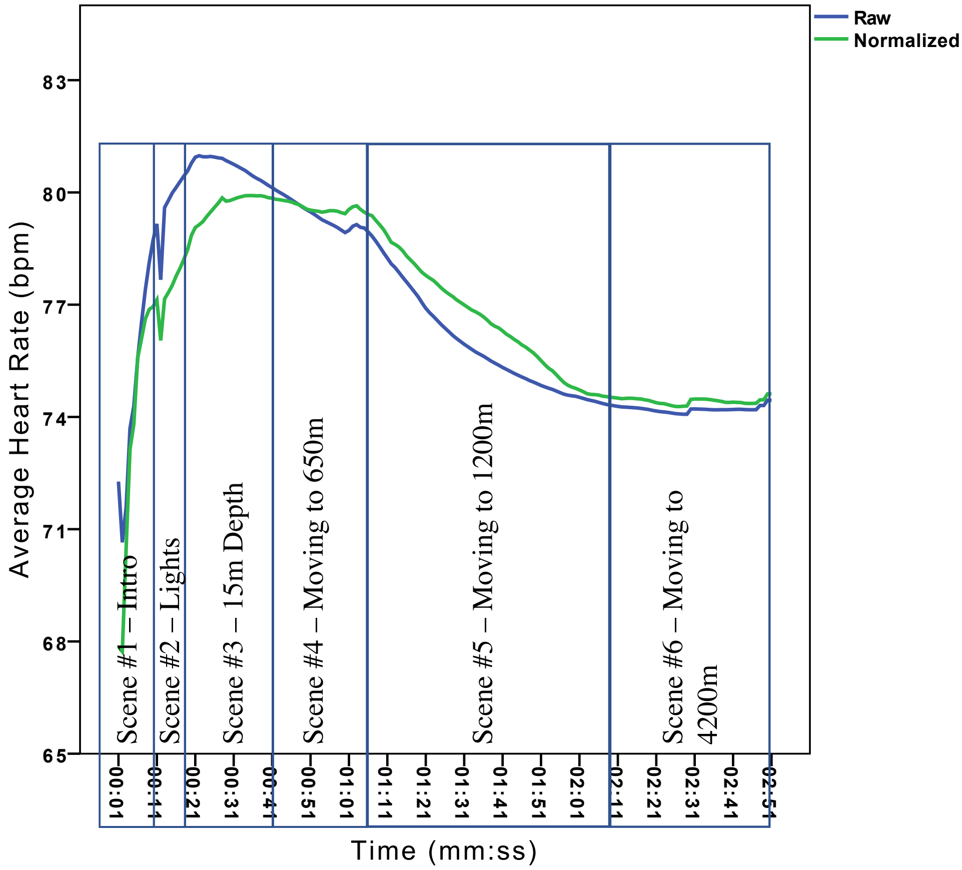


Figure 8.5. Average raw and normalized Heart Rate (HR) during the Aquarium Virtual Reality experience. Participants during Scenes 3 and 4 displayed on average significantly higher HR levels than in all other Scenes.

normality for our HR data ($p = .2$). Then, we grouped our participants' exhibited Heart Rate (HR) by 6 distinct Virtual Reality (VR) Scenes of unequal duration. A Levene's test of homogeneity of variance did not confirm the assumption of equal variances across all the levels of the independent variable Scene ($F(5, 7489) = 22.532$, $p < .001$). For estimating the effect of VR Scene on participants' recorded HR, we performed a one-way repeated measures mixed models analysis of variance (i.e., mixed models repeated measures ANOVA) with participants' HR as a dependent variable and scene number as an independent variable. The analysis displayed a significant main effect of Scene on participants' recorded HR ($F(5, 7451.016) = 88.34$, $p < .001$). Post-hoc pairwise comparisons using the Bonferroni correction displayed significant differences in the estimated means (M_E) of HR for each Scene. Particularly, Scene 3 ("15m Depth":

$M_E = 79.562$, $SE = 1.73$) and Scene 4 ("Moving to 650m": $M_E = 79.575$, $SE = 1.731$) displayed systematically on average the highest Heart Rate (HR) as opposed to all other Scenes, including Scene 1 ("Intro": $M_E = 73.842$, $SE = 1.754$, $p < .001$), Scene 2 ("Lights": $M_E = 77.318$, $SE = 1.788$, $p < .05$), Scene 5 ("Moving to 1200m": $M_E = 76.53$, $SE = 1.719$, $p < .001$), and Scene 6 ("Moving to 4500m": $M_E = 74.462$, $SE = 1.724$, $p < .001$) (see Figure 8.5). Similarly to section 8.3.1, these findings support our initial hypothesis (H1) in that the VR experience will have an impact on participants' physiological responses.

Scene Effect on Heart Rate peaks

Next, we wanted to investigate if the number of HR peaks in participants' exhibited HR levels differed significant on average throughout all VR scenes. For this, we calculated the HR peaks across all participants' HR timelines by annotating a peak event each time an HR was detected. A Pearson's chi-square analysis among peak HR distributions for all Scenes revealed a significant main effect ($\chi^2(5) = 133.428$, $p < .001$, $V = .133$). Particularly, Scene 5 displayed the highest total number of HR peaks ($N = 82$), followed by Scene 4 ($N = 69$), Scene 1 ($N = 64$), Scene 6 ($N = 47$), Scene 3 ($N = 38$), and finally Scene 2 ($N = 10$) (see Figure 8.6). These findings corroborate our initial hypothesis (H1) in that the VR experience will have an impact on participants' physiological responses.

8.3.3 Early Findings

In overall, the results from the two VR studies enable us to answer our previously stated research questions. We were able to unveil significant variations in the average exhibited heart rate levels of our participants, as a result of being exposed to certain VR scenes (H1). In particular, both for the 23 and 46 participants, undergoing the museum and the aquarium VR experience, respectively, VR Scenes that displayed intense movement, such as a train passing and a submarine submerging, were found to increase significantly their average heart levels. In fact, for both VR experiences, the average observed participants' heart rate variations appear similar: a steep increase over the first 2–3 VR Scenes followed by a peak reached at Scenes 2 or 3, only to decrease gradually afterwards (see Figures. 8.1 and 8.5). This is rather interesting in that the same trend appears for innately different VR experiences, which do not even share the same duration (i.e., 5:45 min and 2:51 min, for the museum and aquarium VR experiences, respectively). We speculate that this might also be due to the feeling of excitement induced

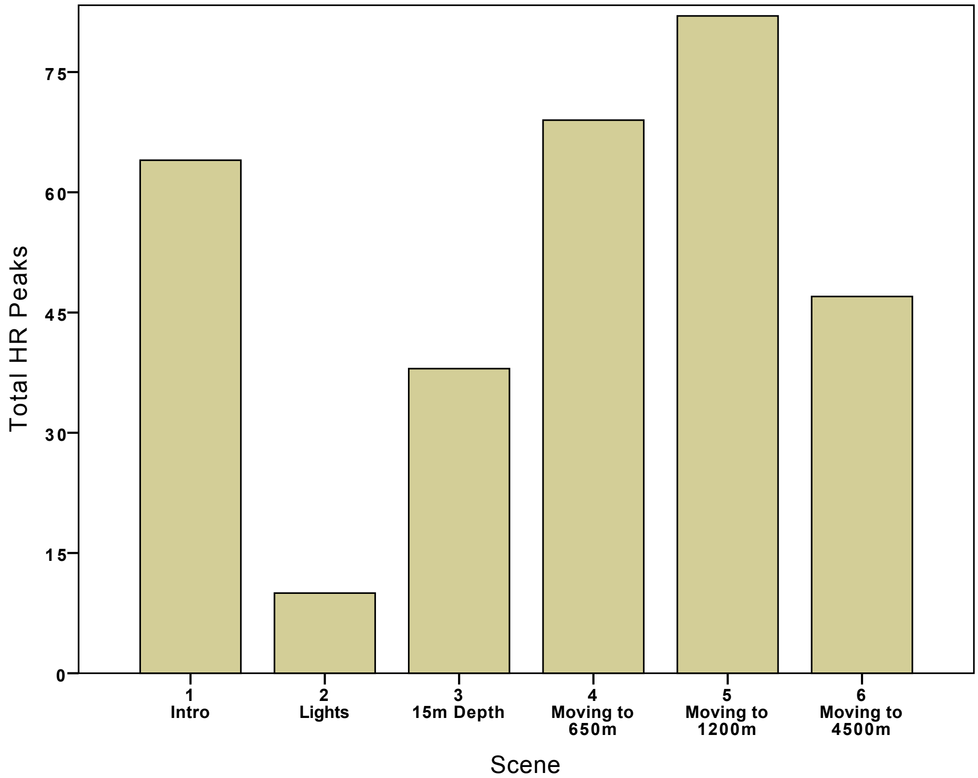


Figure 8.6. Total number of heart rate (HR) peaks per Scene. Scene 5 displayed the maximum amount or HR peaks.

by the immersion in a novel virtual environment and/or the use of novel equipment (i.e., Oculus Rift HMD and Empatica E4 wristband), lasting approximately for the first 30–60 seconds. Despite this plausible "novelty effect", we believe that the observed systematic HR variation is almost entirely attributed to the content of the VR experience per se (i.e., VR Scenes), as previously hypothesized (H1), since participants were fully isolated from external stimuli (e.g., sounds, lighting, etc.).

In fact, the effect of VR Scene was detected in additional physiological features such as HR peaks, at least for the aquarium VR experience. In particular, the aquarium VR Scene 5 ("moving to 1200m") displayed significantly the highest total number of HR peaks, but no effect was observed on the calculated HRV feature. Interestingly, we were able to detect a significant positive correlation between average number of HR peaks per VR Scene and number of recalls per VR Scene right after the VR experience, whereas a significant negative correlation was detected between average number of HR peaks per VR Scene and number of recalls per VR Scene a month after the VR experience. Indeed, this indicates that exhibited physiological responses may play a significant role in what participants remember at later stages of recall (i), as previously hypothesized (H2), possibly highlighting the potential of physiological responses to serve as **memory biomarkers**: predictors of strong or weak memory formation. However, the observed correlations are not particularly strong for drawing safe conclusions yet, and have been detected in a reversed fashion when compared to prior findings in literature. In fact, prior research has shown that high levels of arousal are associated with lower performance in immediate or short-term recall [13, 133, 238]. Similarly, arousal during encoding has been associated with an improved long-term retention and recall performance. In particular, items encoded during high arousal have been recalled better after some time has elapsed, as opposed to shortly after encoding [133, 238], whereas others have found that high arousal, as measured with EDA, benefited both short-term and long-term recall [51, 152]. Nevertheless, all previously cited studies have used either EDA or EEG activity as measures of arousal, in a typical clinical lab, and tap primarily on semantic memory, where for example, participants are asked to recall a set of syllable-numbers pairs or other nonsense learnt items. In fact, Gray's model of emotion theorizes the existence in the mammalian brain of three fundamental emotion systems [87]. The Behavioural Activation System (BAS), a Fight/Flight System (FFS), and the Behavioural Inhibition System (BIS). In particular, the BAS initiates behaviour in response to conditioned stimuli for reward (approach), and has been strongly associated with HR [80]. In contrast, BIS is viewed as an anxiety system that inhibits behaviour in response to cues for punishment, and has been found to strongly associate with EDA [80]. Therefore, it is possible that with measuring HR, and extracted HR features, we detect the activation of an entirely different emotional system (i.e., BAS), and its effect on memory recall, as opposed to measuring EDA (i.e., BIS), and thus observe inverse results to prior findings in literature.

From VR experience design perspective, we wanted to identify those elements that render a VR experience memorable. For this, we extracted 4 different media characteristics namely, *scene format*, *horizon perspective*, *animated elements*, and *sound effects*. So far, we have analysed the influence of scene format (360° photo vs. 3D video) on participants' ability to recall a prior VR experience, and we have found a significant main effect, as previously hypothesized (H3). In fact, VR Scenes in a 360° photo format were recalled significantly more frequently a month after the VR experience, as opposed to right after the VR experience has ended. Interestingly, the same trend is not observed for VR Scenes in a 3D video format. Albeit we have not yet examined the effect of the rest of the VR media characteristics on participants' ability to recall a past VR Scene, and in turn a past VR experience, the results indicate that there is an interplay, at least between the format of a VR scene and memory recall (ii). This is potentially an interesting finding for designing memorable VR experiences, by encompassing scene formats and perspectives that better trigger memory recall.

8.4 Physiologically-Driven Memory Cue Selection

Figure 8.7. PulseCam prototype comprised of a LG G smart watch for continuously measuring one's heart rate, and a Nexus S mobile device for capturing a picture. The Nexus S is situated on one's arm using an armband.

In Chapter 2, we have seen context capture, via lifelogging, emerging as a means for remembering more, mainly through pictures. However, as context capture becomes increasingly mainstream, the volume of captured data also increases, but our capacity for reviewing diminishes. In order to limit picture taking on such devices to only the most memorable moments, we propose a physiologically–driven capture process that adapts the camera capture rate based on one's heart rate. In our prototype — called PulseCam — an Android smartphone worn on the body serves as the picture capture device, adjusting its capture rate based on one's heart rate, as measured by an Android-based smart watch (see Figure 8.7). The purpose of this work is twofold: (a) to examine the potential of PulseCam prototype to capture pictures of significant moments, and (b) to investigate the potential of such pictures to improve one's ability to remember. In this section, we introduce the general approach and describe our yet to be deployed prototype.

8.4.1 Physiological Responses as "Memory Biomarkers"

The idea of using physiological data to distinguish between moments of interest during memory cue selection is not new. In fact, some context-capture approaches rely solely on monitoring physiological data. A type of context-capture device that does not involve pictures is the Affective Diary, a system that records "body memorabilia" (i.e., sensor inputs) from body sensors and mobile media from users' mobile devices. Users of the system were able to remember and reflect on their past using sensor inputs in the form of so-called body memorabilia [213]. Sas et al. proposed AffectCam [200], a wearable system that combines a traditional SenseCam with a Galvanic Skin Response (GSR) wristband for measuring bodily arousal and selecting the most relevant pictures captured. During picture review, an algorithm post-processes the recorded GSR values for identifying arousal peaks and then matches them with the most temporarily adjacent captured pictures. Early results showed that pictures captured during increased arousal greatly improve (> 50 %) one's ability to recall as opposed to pictures of low arousal. The authors also underscore the significance of their approach in reducing the large number of pictures captured via life logging devices such as SenseCam. Drawing on encouraging results from related work and the potential of physiological responses to serve as **memory biomarkers** during VR experience recall, we set forth to examine the prospect of heart rate to drive picture capture. With PulseCam prototype, we are able to monitor one's heart rate continuously, and thus detect in time and in situ any changes in one's excitement levels, that will in turn trigger a picture capture. This provides a strategic opportunity to capture pictures of potentially higher significance, since sampling happens right at the moment of a notable occurrence for the user. In fact, heart rate is highly variable and corresponds greatly to the levels of one's engagement to current experience [190].

8.4.2 PulseCam

The early PulseCam prototype consists of a smart watch[5] and a smartphone (see Figure 8.7). The smart watch is normally strapped to one's wrist and measures continuously one's current heart rate. The smartphone on the other hand, is situated on one's body with its back camera facing forward, for capturing one's field of view. Both the smart watch and the smartphone are paired over Bluetooth Low Emission (BLE). We have developed two Android applications, one for the watch and one for the smartphone, that communicate with each other via Android Wear platform. When heart rate surpasses a certain level, the watch signals a picture capture event to the smartphone. So far, we have been experimenting with a LG G watch for measuring one's heart rate and a Nexus S phone for taking a picture. Our first trials showed that the smart watch battery can withstand up to 8 hours of continuous heart rate monitoring. For now, PulseCam captures a picture every 2 minutes, as a baseline. Every heart rate reading is compared to the moving average of all heart rate readings and if found 10 % higher, a high heart rate picture is captured and labelled. In a future PulseCam deployment, we expect that pictures captured during high heart levels will trigger significantly more vivid recollections when reviewed, and will be considered as more important, as opposed to pictures captured periodically.

8.5 Discussion

Both studies described in this chapter aimed at assessing the interplay between physiological responses and VR scene characteristics with the ability to recall a VR experience at later stages. Hence, the employed VR experiences (i.e., museum and aquarium) serve as a test-bed before moving to the wild and real-life experiences, where we aim to test in realistic settings the potential of physiological responses to act as memory biomarkers. However, before moving "from the lab, out into the fray" there are significant challenges to tackle. First, we were not able to gather sufficient recall responses (only 8 out of 46 participants responded) about the aquarium VR experience at later stage. We attribute this to the ad-hoc nature of our recruitment, as opposed to the more "official" recruitment in the case of the museum VR experience. However, no incentive was provided in both cases. This restricts our ability to generalize our findings when it comes to investigating how our participants' recall ability was affected. Nevertheless, a clear effect of VR scene was detected in the physiological responses

[5] `http://www.lg.com/us/smart-watches/lg-W110-lg-watch-r`

for both the participants of the museum and the aquarium VR studies.

Despite grouping all participants recorded time lines of physiological responses in segments, corresponding to all VR scene durations, a carry-over effect may have manifested across adjacent VR scenes. Nevertheless, the employed mixed-design repeated measures ANOVA, with an "ante-dependence: first order", repeated covariance matrix, accommodates for dependent observations and heterogeneous correlations. In addition, despite using appropriate data preparation and techniques (e.g., outlier detection, normalization, synchronization, etc.) and feature extraction procedures, not all features were successful in unveiling significant fluctuations, and additional data analyses are pending. For example, we have not yet explored at all the effect of participants' profile (e.g., gender and age) on the recorded physiological responses and recall scores. Nevertheless, our findings provide ample ground for further experimentation with physiological responses as **memory biomarkers**, even in everyday life settings. In fact, we expect that the PulseCam application prototype will yield pictures of increased memory value, simply by adapting its capture modality to one's current heart rate levels. In a future version of PulseCam, we will experiment with additional physiological responses (e.g., EDA) for fine-tuning the picture capture process.

8.6 Summary

In this chapter, we presented a set of studies that investigated the potential of physiological responses, measured with Empatica E4, to predict strong or weak memory encoding, and thus driving memory cue selection for drastically improving memory cue quality, at least for VR memories (i.e., RQ4). Albeit not fully-fledged, the analyses showed that, apart from VR scene characteristics, physiological responses and commonly computed features, such as heart rate and heart rate peaks, respectively, may in fact have an impact on one's ability to recall a VR experience at a later stage. Despite these findings are in principle in-line with prior literature, we were surprised to discover an opposite to the status quo trend in the correlation between heart rate peaks and the ability of one to recall a VR experience, both short-term (right after the VR experience) and long-term (a month after the experience). We attribute this effect to the use of heart rate peaks as a measure of arousal, in contrast to the common use of electrodermal (EDA) and brainwave (EEG) activity, and any confounding factors a VR experience might have involved (e.g., novelty effect). Nevertheless, we believe these early findings highlight the untapped potential of physiological responses to serve as "memory biomarkers", greatly enhancing the capture, generation and presentation of memory cues. Thus, drawing on prior literature and on our own early findings, we illustrated PulseCam, a wearable, mobile prototype, that adjusts the process of automatic picture capture based on one's current heart rate levels. Despite not having tested PulseCam yet, we believe it could possibly capture pictures of increased memory value.

In overall, our early findings underscore the aptitude of physiological responses in improving memory cue selection and generation, while highlighting it should not go unnoticed during the design and development of future pervasive memory augmentation systems. In the next chapter, we summarize the totality of contributions for this thesis, we answer our research questions stated in Chapter 1, and we discuss any limitations for such systems.

Part III

Synthesis

Chapter 9

Contribution Summary

In this thesis, we investigated how context captured via ubiquitous technologies can be utilised for augmenting human memory recall, and particularly the recall of episodic and semantic memories. With a series of lab and in-the-wild experiments, we managed to successfully transfer cued recall, an established memory triggering psychological method and the methodological underpinning of this work, from the lab into the wild, for significantly augmenting memory recall in everyday life settings using contextual information. The studies reported in this thesis helped us unveil and even quantify the potential of combining cued recall with modern technologies, while often pointing out future research directions through early application prototypes. In the following sections, we provide a summary of the insights we gained throughout this work and discuss any limitations. But first, we answer the research questions (RQs) stated in Chapter 1 below:

- RQ1 – **Can event-driven mobile capture produce memory cues that augment episodic memory recall?**
  In multiple instances reported in Chapters 4 and 5, we unveiled the potential of event-driven capture to produce memory cues that support mobile cue-based memory recall augmentation. In particular, we showcased that event-driven captured self-face pictures (event-driven selfies) do not only augment memory recall, but also support emotion inference, with an increased effectiveness at later stages (i.e., a week after capture). The significance of this finding lies in demonstrating tangible memory augmentation in the wild, as a proof of concept for our approach. Following up next, we highlighted the potential of (mobile) cue-based memory recall augmentation via event-driven capture in practice, and specifically in the domain of UX evaluation. In particular, we showcased that recalled emotion, as

inferred from event-driven selfies as memory cues, is a good indicator of mobile UX levels for a wide range of mobile interactions. Drawing on these findings and the shortcomings of traditional ecological assessment methods (e.g., ESM), we presented a mobile application prototype that implements cue-based memory recall augmentation, via event-driven memory cue capture, for assessing automotive UX evaluation.

- RQ2 – **How does picture capture modality affect both memory recall and memory cue quality?**
  In Chapter 6, we systematically investigated the effect of manual (both limited and unlimited), and automatic picture capture on memory recall. In particular, we were able to confirm and quantify the so-called "photo-taking impairment" effect [108], and demonstrate that pictures captured manually hold a significantly higher memory value, as opposed to pictures captured automatically with wearable lifelogging cameras.
- RQ3 – **How can the combination of diverse memory cues augment both episodic and semantic memory?**
  In Chapter 7, we applied cue-based memory recall augmentation in work meetings, underscoring the practical potential of our approach in every day life settings such as the workplace. In particular, we demonstrated how episodic memory cues (e.g., lifelogging images) and semantic memory cues (i.e., discussion topics) can synergistically support memory recall, while also establishing a quantifiable baseline for future pervasive memory augmentation systems.
- RQ4 – **Can physiological responses drive memory cue selection for virtual reality memories?**
  In Chapter 8, we brought forward early evidence of a plausible effect of physiological responses on memory recall, when recalling a past VR experience. This suggests that physiological responses are potentially useful in improving the quality of memory cue selection and generation process by driving memory cue capture, though further investigation is required (in real-life context) to reliably support this claim.

9.1 Design Principles for LUIs that Facilitate Memory Recall

As we described in Chapter 2, based on an extensive literature review, we were able to elicit 8 design principles, to which Lifelogging Interfaces (LUIs) should adhere, for effectively supporting memory recall. We briefly present them below:

1. **Expressiveness and meaningfulness of lifelogs.** Utilize the experiential and affective characteristics of lifelogging for increasing the perceived expressiveness and meaningfulness of lifelogs.
2. **Quick overview and prioritization.** Provide a quick overview of relevant episodes, rank them by prioritizing the most relevant, and apply prominent visualization techniques for ensuring they stand out.
3. **Flexible navigation and exploration.** Use entire collections of episodes as retrieval units while supporting flexible navigation and exploration of the corresponding lifelogs.
4. **Adaptation of lifelog content.** Adapt the presentation and visualization of lifelogs to the purpose of recalling, while supporting flexible query generation.
5. **Contextualization of interaction and visualization.** Consider user characteristics and use case application scenarios for contextualizing lifelog visualization and interaction.
6. **Speed and accuracy throttling.** Provide efficient navigation and exploration controls for satisfying speed and accuracy of retrieval on demand.
7. **Multi–modal interaction.** Consider the features of the employed interaction modality for maximizing the capability of LUI in presenting and visualizing lifelogs.
8. **Reinforcement of synergy.** Visualize memory cues for prompting activity, location, and co-presence memory facets that will synergistically support memory recall.

9.2 Event-Driven Selfies Support Episodic Recall at Later Stages

In Chapter 4, we reported a study that utilized event-driven captured self-face pictures (i.e., event-driven selfies), with a dedicated mobile application (i.e., EmoSnaps), for assisting episodic recall at later stages. Our findings show that event-driven captured selfies can indeed jog episodic recall, despite the inherent challenges of highly mobile settings (e.g., blurry or incomplete images, incorrect posture, etc.). In fact, this answers positively our first research question, stated in Chapter 1, on whether event-driven memory cue capture can support episodic memory recall (RQ1). We could think of two ways why this happens: either due to *reconstruction* or due to *recognition*. The reconstruction hypothesis theorizes that participants would use any contextual details included in an automatic selfie (e.g., location, co-presence, etc.) as memory cues for reconstructing the memory of the moment when the selfie was captured, and then recall how they were feeling [189]. Instead, the recognition hypothesis posits that participants would simply recognize their own facial expressions depicted on the selfie, and use them to infer how they were feeling at the moment of capture [192]. At first glance, the results lend credence to the recognition hypothesis, since participants were more able to infer their emotions from an event-driven selfie the longer the time has elapsed since capture. In particular, participants were significantly more accurate in inferring their emotions from an automatic selfie a week after it was captured, than they were at the end of the day. This is a surprising finding simply due to the assumed recency effect [229], where one would expect the exact opposite trend: participants should be better able to recall how they were feeling earlier in the day at the end of the same day, rather than a week later. One possible explanation for this discrepancy could be that the process of inferring emotion from reconstructed episodic memories conflicts with the one of inferring them from facial expressions, and thus disrupting the recall or recognition process.

The emotion inference accuracy achieved by the relevant others may shed additional light on this conundrum. Relevant others were ad-hoc participants recruited only for eliciting emotion from the facial expressions depicted in the event-driven selfies of our main participants. The assumed daily exposure of relevant others to our main participants' facial expressions (being colleagues or significant others) ensures increased capability in accurately inferring their emotions solely by recognition. In fact, event-driven selfies, and the context in which they were captured, were highly irrelevant to significant others, with selfies presented to them in random order and sufficient time since capture for cancelling out any potential memory bias. Therefore, having established a group/condition that employed a rather pristine recognition approach, we can compare it against the rest of the groups/conditions. However, when comparing the significant others' accuracy in inferring emotion from main participants' selfies, to that of main participants a week later, we observe a large difference. Thus, we cannot assume that our participants simply recognized their emotions entirely from their facial expressions when a week had elapsed since capture. Indeed, this logical finding (through abductive reasoning) is in favour of the reconstruction approach, which was probably conflicted by recognition in the early stages of episodic recall (i.e., at the end of the day) with a diminishing influence over time. This reduced influence might be the product of a memory consolidation process (see 2) [7], manifested in the meantime between at least a day after capture and a week later. Therefore, event-driven captured selfies provide effective memory cues for supporting episodic recall, with an increased effectiveness a week after capture [165].

9.3 Event-Driven Selfies for Measuring Mobile and Automotive UX

In Chapter 5, we discussed on how User Experience (UX) is a composite concept, comprised of aspects such as, *user internal state*, *system characteristics*, and *interaction context*, and thus particularly elusive to grasp and measure [105]. However, we are able to approximate UX by measuring its comprising aspects as indicators of overall UX levels. As we have seen, a critical UX aspect is user emotion, and is typically measured with the Experience Sampling Method (ESM), often considered as the "gold standard" of in-situ measurement, since it samples experiences and behaviours right at the moment of their occurrence [142]. Nevertheless, ESM comes with significant drawbacks such as, disrupting current activity, imposing additional reporting burden, and can be inappropriate for certain contexts (e.g., mobile, automotive, etc.). Leveraging on the ESM shortcomings and the success of event-driven selfies to support emotion recall at later stages (see Chapter 4), we proposed an event-driven capture approach of selfies as mediators of user's momentary emotions (i.e., happiness), and indicators of UX. In a subsequent user study, we found that event-driven selfies, captured during a large set of diverse mobile interactions, do not only support episodic recall in general (RQ1), but can also be utilized for successfully evaluating UX of mobile applications. In fact, we were able to unveil significant differences in users' happiness across different kinds of mobile interactions. Interestingly, social interactions such as receive a phone call were associated with reduced levels of happiness, while productivity and social networking applications were associated with increased levels of happiness. Moreover, we found systematic variations of happiness over the course of a day as well as the week, which were largely in agreement with the established findings in positive psychology [165].

As a next step, we wanted to evaluate the effectiveness of our approach in measuring UX in the challenging automotive context. In fact, the uptake of modern in-car technologies and vehicular applications showcase an increased need for in-car UX evaluation. However, ESM is innately restricted, if not dangerous, when it comes to the automotive context, since it can disrupt driver's attention levels. Nevertheless, driver's emotional state per se is an important issue for automotive safety, with a number of driving behaviours being negatively affected by emotions, linking anger or aggression to accidents [139], but also a plausible significant confounding factor for our approach. Therefore, we decided first to investigate how people predict, experience, and recall anger and frustration, before, during, and after commute, respectively, for unveiling any plausible effects of congestion and overall mood that could influence our in-car UX evaluation approach. Our results showed that participants' frustration recall is systematically lower than actual frustration experienced during a traffic congestion. We also found that time of day influences the prediction of anger, being higher in the morning than in the evening as well as, a relation of overall mood with the prediction of frustration [251].

Having established an understanding on how negative emotions exhibited during commute affect the recall of automotive experiences, we designed a mobile application prototype ($eMotion^{+}$) based on LUI principles elucidated in Chapter 2 for retrospectively measuring UX in the automotive context [166]. Instead of ESM, the proposed prototype encompasses a series of so-considered effective memory cues (pictures, video, location, etc.) captured throughout a commute, in a novel LUI, for assisting driver's recall about the commute experience safely at a later stage (e.g., after reaching one's destination). Albeit we did not have the time to test our prototype ($eMotion^{+}$) in an actual user study, based on our findings about event-driven selfies, we believe our approach will accurately measure driver's UX a posteriori. In fact, our results on utilizing event-driven selfies for measuring UX indicate that by increasing the temporal difference between capturing and recall of an experience, we also increase users' ability to infer emotion from their event-driven selfies. The significance of this finding needs to be noted as it suggests that designers, contrary to common sense, should avoid employing related approaches for recent experiences, but rather employ this to recall experiences that lie further in the past.

9.4 The "Photo-Taking Impairment" Effect Exists

In Chapter 6, we reported an extensive user study where we investigated the effect of picture capture modality on memory recall, with and without the support of the captured pictures [164]. Prior work by Henkel et al., has brought forward evidence of the so-called "photo-taking impairment" effect [108], according which the mere act of picture taking can be disruptive for memory formation. Indeed, we were able to confirm the same effect, when comparing free recall performance (i.e., without pictures as memory cues) for participants who captured pictures manually, with free recall performance of participants who captured pictures automatically (with a wearable camera), or captured no pictures at all. Interestingly, this difference in memory performance was more pronounced between participants who used our own capture-limiting intervention, and participants who used no capture modality at all. This possibly indicates that the imposed capture limitation might have been an additional factor that further disrupted the process of strong memory formation (i.e., memory encoding process).

All in all, we are able to answer our second research question stated in Chapter 1 from the perspective of *capture modality*, in that the action of manual capture has a detrimental effect on memory recall at a later stage, when no pictures are provided as memory cues (RQ3). However, we wanted to further investigate why the "photo-taking impairment" effect manifests in a first place. In Chapter 6, we made two assumptions why this effect takes place: either due to distraction caused by manual picture taking (i.e., the status quo), or due to disruption at encoding as a result of having external memory support (i.e., the "Google effect" [212]). However, when having a look at the free recall performance of participants who captured pictures automatically with a wearable camera, both for right after the experience and a week after, we observe no significant differences. This leads credence to the explanation that the "photo-taking impairment" effect is due to the distraction caused by manual picture capture per se.

9.5 Manually Captured Pictures Benefit Significantly Memory Recall

Interestingly, in Chapter 6, we also found that manually captured pictures hold significantly higher memory value than pictures captured automatically with a wearable camera [164]. In fact, a similar trend is mentioned in Chapter 7, where we reported a study that aimed at augmenting the memory recall of past work meetings [169]. In particular, when participants reviewed our memory enhancing intervention, incorporating both topics extracted from a previous meeting (semantic cues), and pictures captured automatically during the meeting (episodic cues), expressed little if any appreciation about the memory value of the pictures per se. Thus, we can answer our second research question stated in Chapter 1 from a *review perspective*, in that manual capture modality produces pictures that are significantly more effective in augmenting one's memory recall about a past experience (RQ2). In fact, in Chapter 6, automatically captured pictures were perceived as holding systematically less memory aid, less ownership, and less semantic gain than pictures captured manually. Interestingly, when we subjected our participants to the task of reviewing their pictures, they spent significantly shorter time in reviewing a significantly larger amount of automatically captured pictures, than they did for reviewing manually captured ones. These findings showcase that periodic automatic capture, in its current state, falls short in producing pictures that can effectively assist remembering.

9.6 Synergy of Episodic and Semantic Memory Cues

Despite the majority of studies, reported in this thesis, focuses primarily on the augmentation of episodic memory per se, we bring forward evidence of augmenting semantic memory too. As such, in Chapter 7, we reported a study where we utilized a slide deck of episodic (pictures) and semantic (keywords) memory cues, as our memory enhancing intervention for augmenting memory recall about a previous meeting. While we did not examine how each cue worked in isolation, informal feedback from our participants confirmed that they mostly relied on the semantic cues (i.e., the extracted topic keywords) when refreshing their memory of a past meeting. We believe this is due to the stationary settings in which our meeting trials took place. For example, a less sedentary meeting setting, such as a seminar, workshop, or an on-site meeting, may increase the memory value of pictures as episodic memory cues.

Moreover, work meetings, particularly goal-driven and in academic settings such as in our study, are inherently considered "semantic" in a way, since they seldom involve emotional or experiential information that in turn is registered in episodic memory. Thus, it was rather expected that in our study semantic cues would be more favoured than episodic ones. However, we believe that both semantic and episodic memory cues can work synergistically (see Chapter 2) for supporting the recollection of a past meeting, or even a past experience, with varying levels of success, depending heavily on the context in which they were captured (RQ3). Therefore, LUIs that encompass hybrid memory cues maximize their potential in augmenting memory recall, such as the $eMotion^{+}$ mobile application presented in Chapter 5 (e.g., speed and time as semantic cues, whereas video of facial expressions as episodic cues) [166].

9.7 Physiological Responses as "Memory Biomarkers"

In Chapter 2, and elsewhere in this thesis, we have stressed the need for "selectivity, not total capture" first introduced by Sellen and Whittaker [210], when it comes to lifelogging for augmenting memory recall. In fact, selectivity was the main idea behind the MGOK mobile application prototype [167], which was employed as a restrictive capture intervention during the campus tour experiment, drawing on the paradigm of the old film cameras. (see Chapter 6) [164]. One of the main hypotheses was that the imposed capture scarcity would induce a higher feeling of selectivity to our participants, that in turn would lead them to capturing pictures of high importance, and thus of increased memory value. Nevertheless, the results showed that participants were rather puzzled by this uncommon capture restriction, perhaps at the cost of expending additional attention reserves necessary for strong memory formation (i.e., memory encoding), and thus exacerbating the "photo-taking impairment" effect. However, when we later examined the memory value of pictures captured during an artificially-induced, highly-selective capture behaviour, and while controlling for number of pictures captured, we were surprised to find out that selectivity did produce pictures of significantly increased memory value, as opposed to those captured automatically (RQ2) [164]. These findings suggest that the "selectivity, not total capture" design principle may need to be implemented in a system level, rather than imposed in a user level.

In Chapter 8, we describe an alternative strand for achieving selectivity in lifelogging for augmenting memory recall; by utilizing physiological responses as "memory biomarkers", indicators of weak or strong memory encoding. Prior work has shown that physiological responses, as indicators of arousal, relate to memory formation and recall, and have even been used for filtering lifelogging content (e.g., "AffectCam" by Sas et al. [200]). Drawing on these findings, we also wanted to inquire into the potential of additional physiological responses (e.g., heart rate) to improve the selection of memory cues but in a more systematic way. We decided to measure the impact of physiological responses on the recall of Virtual Reality (VR) experiences. The selected VR experiences are well-structured, highly-controlled (no external stimuli), and homogeneously encountered, ensuring that all participants are subjected to exactly the same stimuli, thus providing ample and ideal ground for testing our hypotheses. Indeed, we were able to detect significant correlations between participants' physiological responses recorded during particular scenes of the VR experience, and participants' ability to recall these scenes at later stages. However, the observed correlations appear in a reverse fashion to what is reported in literature (i.e., negative for short-term recall and positive for long-term) [13, 133, 238], an effect we believe is largely attributed to investigating heart rate (HR) peaks in relation to recall, as opposed to typically using electro-dermal activity (EDA) in literature. However, HR has been found more sensitive to visual stimuli, as opposed to EDA, with an average response rate of approximately 2 seconds (depending on stimulus valence) [175], as opposed to an approximate response of 5 sec for EDA [88]. In fact, EDA has been prevalently used for gauging arousal, and has been linked with the intervention of behavioural inhibition system (BIS), responsible for inhibiting behaviour, whereas HR has been connected with the behavioural activation system (BAS), known for initiating behaviour [80].

Albeit early and in the sedentary context of VR, our results suggest that physiological responses may be good indicators of memory formation (i.e., "memory biomarkers"), and hence could be used for improving or even driving memory cue selection process (RQ4) during lifelogging in a system level. A subsequent deployment of the "PulseCam" prototype [168] will more reliably assess the potential of physiological responses in augmenting memory recall in everyday life settings.

9.8 Limitations

Throughout the course of our research, we have conducted both lab and field experiments for testing the effectiveness of our interventions in augmenting memory recall. In fact, memory recall is a concept particularly notorious to reliably gauge, as it appears to fluctuate based on a plethora of factors one is not able to fully control or even predict. Such factors can often significantly influence our participants' memory recall ability, being either, but not limited to, intrinsic (perception, physiological, and emotional state, etc.), or extrinsic (current context). For example, simply a night of bad sleep could hinder one's concentration and attention levels during a subsequent experiment, and thus influence one's overall ability to encode new memories or recall old ones. Typically, our experimental designs cancel out most of such effects, while we recruit a minimum number of 12 participants that are tested for sufficient time periods and for multiple times over the course of the study, before drawing any conclusions. For example, in Chapter 7, where we presented our memory-enhancing intervention for work meetings, we measured participants' memory recall twice per condition over the course of a month, ensuring a more robust effectiveness assessment for our memory intervention. Nevertheless, often unpredictable cognitive and social biases, as well as known physiological effects, might have affected our results in expected and unexpected ways. Below, we describe some of these cases.

9.8.1 Social and Cognitive Biases

In Chapter 5, we described a field study where our participants used a mobile application (EmoSnaps) for the event-driven capture of self-face pictures (event-driven selfies) at predefined mobile interaction moments (e.g., screen unlock, SMS sent, etc.) [165]. In this study, our assumption was that our participants would either recognize or recall their emotions from the captured selfie, and use their emotion ratings as a proxy for assessing UX. However, when inquiring into the rationale of our participants when rating their event-driven selfies, we were surprised to discover that female participants were particularly concerned with the aesthetics of their selfies, and reported that when they were insufficiently appealing they discarded them. Clearly, this is a case of a social bias we could not have predicted a priori, and may have even influenced the way female participants have rated their event-driven selfies in overall.

Other factors that may have affected the quality of our results are cognitive biases and particularly memory biases. In Chapter 6, we presented a study that investigated the effect of picture capture modality on memory recall in a campus tour [164]. All of our participants were University students of varying familiarity with the campus whereabouts. However, we proceeded to assessing our participants' ability to recall the campus locations and the items the visited without controlling for their prior experience with campus. Nevertheless, we expect a rather small impact on our results, since the displayed items and the corresponding details were rather obscured, even for a graduate student to know. In Chapter 8, we described a study that entailed the presentation of VR experience about the Swiss province of Ticino, and the ability of our participants to recall it at later stages. In this case, we opted out of recruiting participants from Ticino by conducting our experiment in Zurich, in an effort to reduce familiarity and any memory biases. However, we were not able to recruit participants that had never been to Ticino before, since it is a particularly popular vacation destination in Switzerland. Despite the manifestation of memory biases in the studies reported, we can assume that they are almost evenly distributed across conditions, and hence their impact on our results is rather minimized.

Typically, we take extra effort to minimize memory biases by counterbalancing conditions or randomizing how our intervention is delivered. For example, in Chapter 3, where we investigated how much relevant others were able to recognize the emotions from event-driven selfies of our main participants, we took extra care to randomize the event-driven selfies displayed while waiting for at least a week to elapse since capture. Hence, in this way we minimize any memory bias that relevant others may have from recalling a common experience with our main participants. However, in other cases we are not that successful in removing or counterbalancing memory biases, simply because the employed study design dictates that certain aspects are measured in a certain order. For example, in the study about augmenting memory recall of past work meetings (see Chapter 7) [169], we wanted to inquire into which extent our participants thought that our memory intervention helped them have an effective meeting. Thus, we subjected our participants to our memory intervention before a meeting, and we measured their perceived effectiveness after the meeting. However, we were surprised to discover no significant difference in participants' perceived effectiveness between using and not using our memory intervention. We believe this happened due to administering our questionnaires after significant time had elapsed since our memory intervention, and thus participants had forgotten about it, or simply were biased by the outcome of their own meeting.

Finally, same as with every experiment in human subject research, we expect a certain degree of observation bias to manifest in our studies. Thus, it is possible that our participants' performance when using our interventions may have been affected from the very fact that a researcher was nearby or from the use of measuring equipment (e.g., E4 wristband). Also, by adhering to sound research ethics, we always informed our participants about the purpose of our experiments, and thus this might have in turn influenced their behaviour, or performance. Fortunately, with carefully tailored study designs and knowledge on the biases introduced by our methodological modus operandi, we can avoid or at least anticipate and quantify the impact of biases on our results [183].

9.8.2 Other Known Effects

Apart from the previously described, often unpredictable biases, there are often unavoidable effects known to manifest under certain conditions, and have to be taken into account when analysing and interpreting our results. In particular, when one deploys prototypes that are tested in everyday life settings, one should be aware of known physiological phenomena imposed by participants' circadian rhythm and their daily schedule (e.g., fatigue). In fact, when it comes to long deployment periods, that may span from a week to longer than a month, the impact of these effects on our study results may be immense. In Chapter 5, where we presented the study for UX evaluation with event-driven selfies [165], we encountered known effects imposed by daily life on the way our participants rated their past emotions based on their event-driven selfies. For example, we unveiled

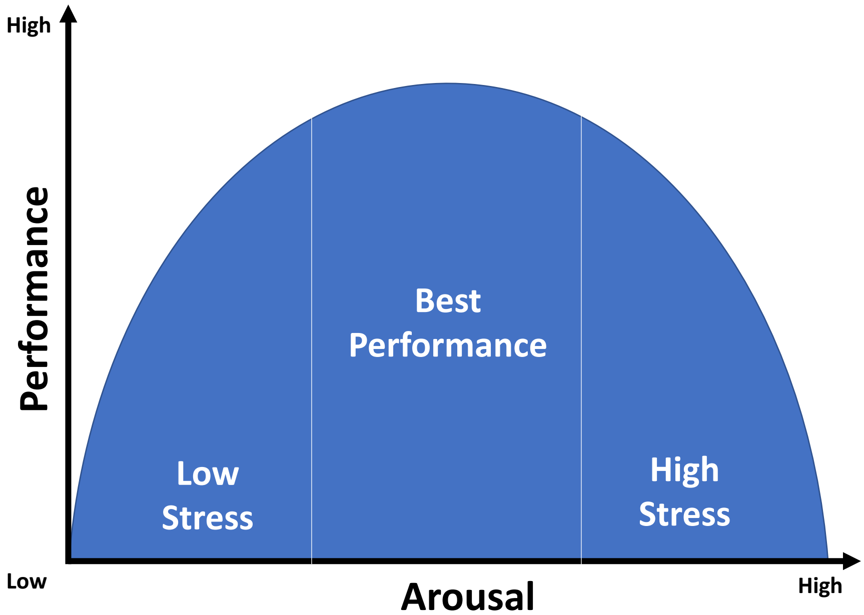


Figure 9.1. The Yerkes-Dodson curve theorizes a certain threshold of arousal contribution to (recall) performance. Past that threshold, the arousal contribution becomes negative.

systematically lower happiness ratings for morning hours, Mondays ("blue Monday" effect), and a higher sefie discard rate at night. In contrast, happiness was systemically rated higher on Sundays ("weekend" effect).

When it comes to human memory research, there are certain effects known to manifest during memory experiments, and we were able to discover too. For example, when recalling a (negative) past experience, our recollected emotions are generally positively–skewed, a phenomenon known as the "rosy view effect" [161]. We were able to unveil this effect in a study briefly reported in Chapter 5, where we report results on participants' recalled frustration, as opposed to actual frustration during commute [251]. Other known memory effects are the primacy and recency effects, according which participants recall better the starting and the ending point of an experience or series of memorized items [162]. In Chapter 6, we describe how we encountered these effects during the campus tour experiment, and how our participants often changed their capture behaviour to compensate for such effects, particularly when under the experimental condition of the artificially imposed capture limitation [164, 167]. Similarly, in Chapter 8, where we describe the study about the recall of the museum VR experience [153, 154], we were also able to unveil primacy and recency effects on which VR scenes participants could recall right after and a month after the VR experience (see Figures 8.2 and 8.3).

Novelty is perhaps another effect that we need to take into account when interpreting our results, particularly for studies that involved the use of "exotic" equipment or a novel experience. We believe the novelty effect was markedly present during both independent VR studies reported in Chapter 8. In fact, in both VR experiences we were able to detect a rather steep increase in participants' average heart rate levels during the first 30–60 seconds. We attribute this phenomenon to the novelty effect imposed by encountering a novel VR experience, but also the use of yet uncommon equipment (e.g., Oculus Rift). Therefore, the novelty effect may have influenced our results and how we interpreted them, since arousal was of key importance in the effort of unveiling any effect on subsequent experience recall. Nevertheless, despite the effect of novelty on arousal, we consider that the arousal–recall relationship follows the Yerkes-Dodson empirical law, drawn as a "∩" curve (see Figure 9.1).

9.8.3 Context is Everything

In this thesis we have utilized context in terms of captured contextual information for generating effective memory cues that substantially aid one remember a past experience (episodic) or obtained knowledge (semantic). However, context is also of paramount importance when it characterizes the setting for which a memory intervention is intended. Throughout this work, we have presented studies that enabled established memory augmentation techniques (i.e., cued recall) to escape the narrow boundaries of a typical Psychology lab, and find their way into everyday life settings. However, the effectiveness and the applicability of our interventions and the ecological validity of our results are questionable when the context is not the intended. For example, in Chapter 8, we investigated the effect of physiological responses on participants' ability to later recall a VR experience, drawing the conclusion that physiological responses may be able to serve as "memory biomarkers". Evidently, the sedentary context of VR is entirely different than the versatile context of everyday life, where additional factors should be taken into consideration (e.g., mood, physical activity, etc.).

In other cases, our experimental intervention may be considered promising conceptually and even experimentally, but rather unfit for the modern technological context of nowadays. For example, in Chapter 6, we investigated the effect of capture modality on memory recall by introducing the skeuomorphic MGOK mobile application that artificially limits the amount of pictures one can capture [164, 167]. As we have seen, the assumption behind MGOK is that the imposed capture scarcity would lead to capturing pictures of higher memory value, and at the same time ameliorate the "photo-taking impairment" effect [108]. However, inviting participants to be more selective during capture did not have the expected results of mitigating the "photo-taking impairment" effect and producing pictures of suggestively higher memory value. Nevertheless, even if the results were positive, transferring such a capture limiting intervention into the unlimited storage capacity of the today's world, would have found little if any acceptance at all. Therefore, the context in which our interventions were tested and applied, heavily influences the generalizability of our findings.

Chapter 10

Conclusion and Future Work

In this thesis we described our viewpoint on how established ubiquitous technologies can benefit human memory recall. Our modus operandi postulates the utilization of contextual information derived from a plethora of modern technologies (e.g., lifelogging) for generating **memory cues**, hints displayed over time for augmenting one's memory recall about a past experience (episodic memory) or prior knowledge (semantic memory). In overall, findings from field deployments and lab studies were in favour of (mobile) cue-based memory augmentation, underscoring the potential of event-driven captured self-face pictures (event-driven) to serve as effective memory cues for augmenting emotion recall, even in the challenging mobile context. Notably, we found that the effectiveness of event-driven as memory cues increases after substantial time has elapsed since capture (i.e., a week), and we showed that they could be utilized for UX evaluation in diverse contexts. We were also able to gauge the value of typical pictures in assisting memory recall, unveiling a higher memory recall augmentation potential for those captured manually, as opposed to pictures captured automatically. This possibly explains why current lifelogging approaches that utilize wearable cameras are rather limited in substantially assisting memory recall. Indeed, lifelogging is a great source of memory cues, but the design of Lifelogging User Interfaces (LUIs) is integral as it should adhere to specific design principles, which we have outlined in this thesis, for effectively assisting remembering. We consider the reinforcement of synergy between organic memory and the LUI as a critical design principle, and we showed it can be maximized by the incorporation of memory cues that aid different types of memory. In fact, we draw on the combined power of hybrid memory cues that jog concurrently episodic and semantic memory, while establishing a quantifiable baseline for memory augmentation in the workplace. Finally, we brought forward nascent evidence that physiological responses could improve the process of memory cue selection (in VR memory recall), clinging to the dogma of "*selection, not total capture*" for lifelogging that supports memory recall [210].

In the next section, we discuss any implications concerning the plausible broad adoption of pervasive memory augmentation systems in the future. We then present the future directions of this work, featuring approaches and systems that will holistically amplify human cognition, and not just human memory, by drastically improving the way we interact with technology. We conclude this thesis by reporting our final remarks.

10.1 Implications

Quite certainly, the design of future pervasive memory augmentation systems will have to face challenges that span from the design of a secure system architecture and novel information retrieval practices, to the design of user interfaces that adhere to sound design principles, some of which we have elicited and reported in Chapter 2 (see Section 2.2.4). Even so, taking into account the immense advancements in the field of big data and artificial intelligence, we believe the greatest challenges lying ahead are social and ethical rather than technical. Perhaps the most recent and strikingly accurate take on such a future memory augmentation system, and its hypothetical social and ethical implications, has been presented in the popular "*Black Mirror*",[1] a dystopian, Sci-Fi drama TV series enacted in the short-term future, in the form of independent episodes. In particular, the third episode of the first season (2011) is titled as "*The Entire History of You*",[2] and envisions a future where widespread memory implants record and provide instant access to precise archives of our entire life, enabling one to remember everything and forever. Despite we do not fully align with the bleak prospects described in the series, we do believe a memory augmentation system entails serious social and ethical challenges, some of which we discuss next.

First and foremost, the security and privacy implications regarding the use of pervasive memory augmentation systems have been thoroughly discussed by Davies et al., with a particular emphasis on how one's "digital memories" can be protected from theft, manipulation, or deletion, while remaining accessible to those who matter (e.g., family or mankind) [56]. However, even if one's digital memories are fully protected and fully accessible for a lifetime and beyond, there are certain implications that stem from the use of memory augmentation

[1] http://www.imdb.com/title/tt2085059/?ref_=nv_sr_1
[2] http://www.imdb.com/title/tt2089050/?ref_=ttep_ep3

technologies per se. As such, "Recall-Induced Forgetting" (RIF) describes the phenomenon during which enhancing specific memories comes at the cost of attenuating others [1]. In fact, the frequent recall of specific memories through techniques such as cued recall, improves the probability of spontaneously retrieving reviewed memories in the future, but also reduces the probability of recalling related but non-reviewed memories [56]. For example, augmenting one's memory solely about a beautiful day at the beach may attenuate the memory of what a hassle was to get there in a first place. Even the frequent retrieval of so-called "targeted memories" (i.e., memories subjected to augmentation) may obfuscate them, by rendering them into a malleable state of re-consolidation, during which targeted memories are more prone to alteration. This effect is known to manifest during the recall of autobiographical memories, such as the ones that this work strives to enhance [18, 121]. Further research is needed to fully fathom the long-term impact of such memory augmentation systems on human memory, and thus one should be aware of such plausible effects when employing similar memory-enhancing technologies.

Perhaps a second look at the elicited design principles for LUIs is enough for one to realize that many of them do not only concern the design of LUIs, but could also describe best practices when it comes to the use of future pervasive memory augmentation systems. "*Synergy, not substitution*" is maybe the best example of a design principle and a best practice simultaneously, as it stresses the synergistic purpose that memory augmentation systems should serve. In particular, such a system should never strive for replacing our organic memory, but rather work in tandem with it by employing adaptive memory cuing techniques (i.e., adapt its support to one's actual recall needs). A future scenario where a memory augmentation system acts antagonistically towards our organic memory, could bear largely unforeseen consequences, quickly turning out to a "*Black Mirror*" dystopia. Presently, perhaps the most resembling analogy is the so-called "Google effect", according which nowadays there is an ever-decreasing need and will to memorize actual knowledge (i.e., by tapping our semantic memory), and even skills, since knowledge is literally at our fingertips via modern technologies (e.g., a smartphone) [212]. Instead, one only needs to memorize and develop, knowledge retrieval patterns and sharp querying skills, respectively (e.g., inputting the right keywords in Google search engine). Therefore, our intention with this work is to initially use technology for augmenting one's memory about a past experience or obtained knowledge, and gradually reduce the level of our intervention to the extent one would only rely on one's memory. Ideally, the memory augmentation achieved should persist even in the absence of our intervention, for minimizing increased reliance and any plausible detrimental effects on one's organic memory.

Finally, throughout this thesis, we have entirely assumed that our memory-enhancing approaches are purposed for the able-bodied individuals, and hence we have only trialled our interventions with healthy participants. Nevertheless, the impact of our work could be further increased when reaching out for targeted audience in dire need of memory improvement. We believe patients suffering from amnesia, or other causes of memory deficits, would be benefited the most from our work. In fact, the ability to form new memories is known to aggravate with ageing, often as a result of chronic neurodegenerative diseases and persistent disorders such as Alzheimer's and Dementia, respectively. However, the actuation of modern technologies for facilitating memory recall, as proposed in this thesis, bounds the prospects of our approach to the technological shrewdness of the individual. Therefore, the actual challenge lies in envisioning efficient solutions, adapt policies, and better target services and products that solve societal problems such as the technological age gap.

10.2 Future Work

With do-it-yourself (DIY) transcranial direct-current stimulation (tDCS)[3] [224] and DIY CRISPR-Cas9 DNA editing[4], advanced brain implants[5] [110], and other prosthetic technologies[6] and general purpose AI being ante portas of our daily lives, one reasonably wonders what is the next "Big Thing". As we have discussed (see Chapter 2), we believe that the next big step in HCI is the attempt to bring closer together the user (human) and the system (machine), narrowing the so-called "cognitive gap", the lack of knowledge about the user's cognitive states from a system perspective. Such an advancement will not only bear great benefits for one's memory (e.g., by readily interacting with LUIs), but could potentially revolutionize the entire spectrum of our cognition (attention, decision-making, learning, etc.). In the following sections, we briefly present our envisioned architecture for bringing closer the human and the machine, by bridging the so-called cognitive gap [170].

[3] http://www.bbc.com/news/health-27343047

[4] https://www.scientificamerican.com/article/mail-order-crispr-kits-allow-absolutely-anyone-to-hack-dna/

[5] https://www.newscientist.com/article/2153034-brain-implant-boosts-human-memory-by-mimicking-how-we-learn/

[6] https://futurism.com/mind-controlled-robotic-arm-johnny-matheny/

10.2.1 The Cognitive Information Layer

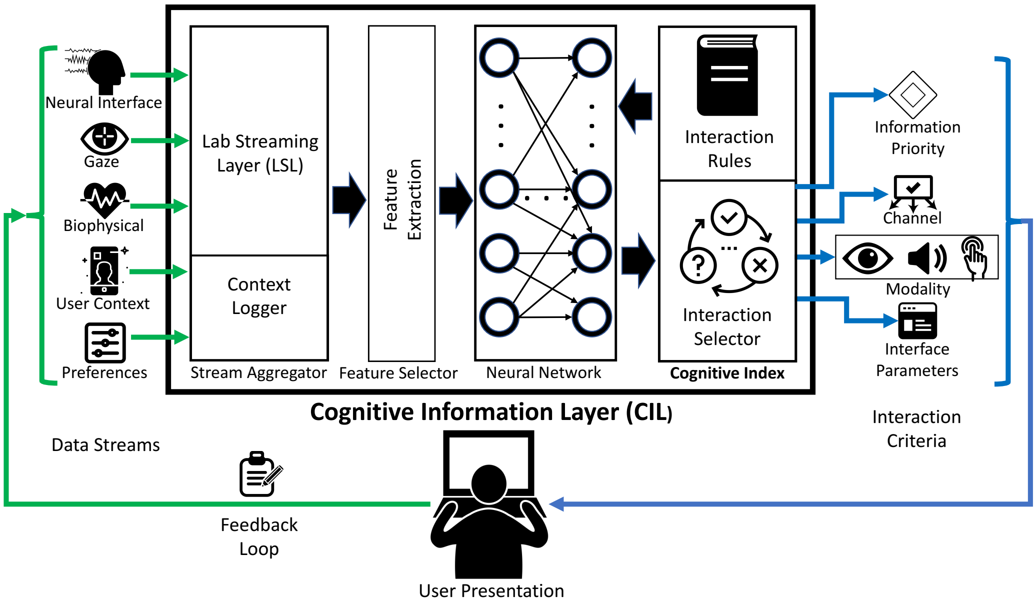


Figure 10.1. The envisioned Cognitive Index Layer (CIL), receiving as input multi-modal data streams and outputting adequate interaction criteria for adjusting user presentation to user's current cognitive state.

The main component of the envisioned architecture is an intermediate layer between the human and the machine, responsible for continuously monitoring the user's cognitive and affective states and informing the machine side for performing adequate interaction interventions. As "machine", we consider systems with which we interact frequently and increasingly depend on, such as personal computers, smartphones, smartwatches, public/personal displays, and soon any IoT device. Interaction interventions regulate the information flow between human and machine to match the current cognitive and affective state of the user, user context and task at hand. An overview of the proposed cognitive information layer is shown in Figure 10.1. In short, the envisioned architecture is based on real time multi-modal data stream inputs that reveal both current cognitive and affective states, task at hand, and user context. Next, features are extracted and classification takes place for selecting the adequate interaction criteria based on a constantly updated "cognitive index". Finally, the user presentation is adapted on the fly according to the cognition-aware interaction criteria outputted. In the next subsections, we describe in more detail the primary stages and components of the envisioned architecture.

Multi-modal Data Streams

We are currently experiencing an upheaval of data that characterize one's physiological states, mainly due to the miniaturization and reducing cost of sophisticated wearable physiological hardware [237]. Mobile eye-trackers (e.g., Tobii Pro Glasses 2[7]), affordable portable neural interfaces (e.g., OpenBCI Ultracortex "Mark IV"[8]), and physiological monitoring wristbands (e.g., Empatica E4), produce a sheer volume of physiological data. Gaze behaviour, Electroencephalography (EEG), Heart Rate Variability (HRV), and Electro-dermal Activity (EDA) are just a few examples of physiological data that can reveal one's cognitive or affective states, stress levels and other intrinsic information [22]. This plethora of cognition-descriptive information can be utilized in conjunction with user context, task at hand, and preferences, to greatly increase the accuracy at which user cognitive and affective capacities can be estimated [16]. As shown in Figure 10.1, all multi-modal data streams are aggregated continuously and then pushed into the feature selection stage, where data is filtered and useful features are extracted. Feature extraction aims at describing the acquired data streams with as few relevant values as possible. Such features should capture the information embedded in those data streams that is relevant to describe the mental states, while rejecting the noise and other non-relevant information. The next step, denoted as "classification" assigns a class to a set of features extracted from the data streams. This class corresponds to the kind of mental state identified. Classification algorithms are known as "classifiers". Typically, for learning which kind of feature vector corresponds to which class (or mental state), classifiers try either to model which area of the feature space is covered by the training feature vectors from each class, or they try to model the boundary between the areas covered by the training feature vectors of each class mostly used in BCIs.

The Cognitive Index

When first installed, the Cognitive Index (CI) is simply a catalog of appropriate interaction rules for efficient management of one's cognitive resources, extracted from guidelines and best practices available in literature. However, the CI is constantly updated to match the cognitive capacities, preferences, task at hand and context of each individual user. Next, the CI is responsible for selecting the appropriate interaction criteria in terms of information prioritization (e.g., e-mails over social media notifications when at work), adequate channel (e.g.,

[7] https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/
[8] https://shop.openbci.com/products/ultracortex-mark-iv/

mobile vs. stationary), modality (e.g., visual, audible, or tactile), and certain interface parameters (e.g., font size), as shown in Figure 10.1. Ultimately, the CI will be enriched by interconnecting the CIs of different users for creating a universal pool of cognitive interaction mapping.

The Feedback Loop

Finally, the outputted interaction criteria have resulted in tangible interaction interventions that are delivered in real time to the user, by appropriately altering user presentation to better match current cognitive capacities and needs. At this point, a closed feedback loop informs in situ the cognitive information layer about the effect that the modified user presentation has on user's cognitive state via behavioural measures (e.g., task completion times, motion patterns, etc.). This serves as a real-time assessment mechanism for evaluating the effect of the selected interaction intervention, and adapting the interaction criteria accordingly.

10.2.2 The Cognitive Application Framework

Knowledge obtained from the deployment of the CIL will be the basis for eliciting requirements and guidelines for informing the design and development of cognition-aware applications. "Cognitive applications" will consider user cognitive states for adapting information flow, interaction techniques and interface parameters. Design guidelines, requirements and best practices will be incorporated into a cognitive application framework, along with a dedicated cognitive Application Programming Interface (API). The cognitive API will connect cognition-aware applications with the CIL for providing interaction criteria that facilitate communication between human and machine. For instance, the cognitive API will inform cognition-aware applications such as a text editor, an (mobile) e-mail client, and a music player, that user's attention levels are currently scarce. In turn, the cognition-aware text editor will increase window and font size, and contrast, for keeping user focused to a text typing task. Simultaneously, the cognition-aware e-mail application will suppress incoming e-mail notifications for avoiding disrupting user's attention, and the cognition-aware music player application will opt in for music that helps one focus. This application scenario showcases that the CIL, via a dedicated cognitive API, could synergistically amplify human cognition by informing and adapting information flow and interaction across multiple applications and channels (e.g., mobile or stationary) simultaneously and surreptitiously. This CIL characteristic may be vital for our cognition, rectifying and regulating the fierce competition of modern applications over our limited cognitive capacities.

10.2.3 Challenges

Clearly, the proposed approach is difficult to realize. From a software architecture perspective, challenges lie throughout the entire stack of the envisioned cognitive information layer. First, multi-modal data streams can be largely heterogeneous with wildly varying sampling rates and fundamentally divergent acquisition logic. For example, an EEG signal is continuously obtained from 250–500 Hz, whereas a location transition can be sampled asynchronously, yet both data types need to be synchronized and processed together. Even so, accurately inferring users actual cognitive state (e.g., for ground truth acquisition) remains a conundrum. Selecting the right classifier for features derived from a plethora of multi-modal data streams, could be a solution but also a considerable challenge. In Figure 10.1, we assume a neural network (NN) as the most adequate classifier candidate due to the NN's ability to deal with highly heterogeneous input. Finally, the outputted interaction criteria should be implemented within the cognitive application per se. This means that (mobile) applications that one uses daily (e.g., an Internet browser) will need to be re-programmed to accommodate for changes that the outputted interaction criteria recommend. However, we believe that the biggest challenge is a shift in how designers and developers create software for everyday use. Cognition-aware software will have to consider the users cognitive states and adapt its functionalities and presentation accordingly.

Beyond the technical challenges, we also foresee more practical challenges. From a sociotechnical perspective, challenges lie primarily in user adoption. Physiological sensing hardware (e.g., an EEG system) is still expensive, cumbersome to use, tiresome to wear for extended periods of time, and socially unacceptable. Furthermore, current hardware solutions are prone to noise generated from movement or muscle activity, heavily degrading the quality of the acquired signal. However, as hardware miniaturization progresses, with wearables and IoT becoming more prevalent, we expect that physiological sensing will become gradually mainstream, perhaps even in an implant level (e.g., Neuralink). From a user privacy perspective, the streams of data supplied to the CIL are highly sensitive, able to reveal apart from user's cognitive state, also health and affective states, and thus data protection measures and policies will be of paramount importance. Also, the diffusion of knowledge about our cognitive states to a network of interconnected objects (IoT) may also raise unexpected ethical and security concerns [28]. Despite the considerable challenges, we believe that the proposed architecture, or a similar one, will be implemented and seamlessly integrated in an operating system level in the mid-term future. This integration will form and enforce design and development policies for creating cognition-aware applications.

10.2.4 Summary

Human brain evolution is thought to have already reached its apex together with our cognitive capacities. Some argue that one way forward is through achieving human-machine symbiosis, the notion of human converging with the machine. While sounding like a Sci-Fi scenario, we argue that not only could it soon be a reality, but also a necessity. We introduced the cognitive information layer as an inset between human and machine, for informing the machine side about current user's cognitive state and facilitating human-machine interaction. Thus, we define human cognition amplification as the optimization of existing cognitive resources, rather than extending human abilities beyond the humanly possible. We illustrated how by supplying a set of multi-modal data streams to the proposed layer, it can output a set of interaction criteria as a pivot for manipulating user presentation with cognition-aware applications. We identified a range of challenges, including the need to reform traditional software (and not only) design thinking, so to create software that respects our cognitive capacities. We believe that the human brain and technology can and should be able to work more closely in tandem for amplifying our cognitive capacities in the era of distractions and information overload.

10.3 Final Remarks

Modern technology and contemporary trends such as lifelogging and the quantified-self movement, allow us to capture personal contexts unobtrusively and surreptitiously. Personal mobile devices such as laptops, smartphones, digital and wearable cameras (e.g., Narrative Clip), wristbands (e.g., Fitbit), or smart watches can all record — both implicitly and explicitly — a wide range of contextual information about one's daily life. In this thesis, we investigated the potential of personal contextual information, in the form of memory cues, to support the recall of past experiences and/or obtained knowledge. By leveraging on the latest technological advancements in the field of ubiquitous computing, we showcased how an established psychological method, such as cued recall, can be successfully transferred into the wild for yielding tangible memory improvement in the challenging settings of everyday life. In fact, we significantly and measurably augmented our participants' recall capacity in a diverse set of contexts that surpass the narrow boundaries of a typical Psychology lab, enabling us to draw valuable insights for the design and development of future pervasive memory augmentation systems. We also showcased that any memory augmentation system should strive for collaborating with our organic memory instead of replacing it, and ideally should bear a memory augmentation effect that will persist even in the absence of the system per se. We thus believe that we have contributed, with a small but considerable fraction, to the fundamental understanding of human memory and its interplay with today's modern technologies. Above all, in this thesis we shared our vision of a world where memory cues are ambiently displayed for drastically transforming how humans manage their memories, creating new avenues and ample ground for novel application scenarios and future research, respectively.

Bibliography

[1] Anderson, M. C., Bjork, E. L. and Bjork, R. A. [2000]. Retrieval-induced forgetting: Evidence for a recall-specific mechanism, *Psychonomic bulletin & review* **7**(3): 522–530.

[2] Anderson, M. C., Bjork, R. A. and Bjork, E. L. [1994]. Remembering can cause forgetting: retrieval dynamics in long-term memory., *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**(5): 1063.

[3] Anderson, S. J. and Conway, M. A. [1993]. Investigating the structure of autobiographical memories., *Journal of Experimental Psychology: Learning, Memory, and Cognition* **19**(5): 1178.

[4] Annas, G. J. and Grodin, M. A. [2008]. The nuremberg code, *The Oxford Textbook of Clinical Research Ethics*, Oxford University Press New York, pp. 136–140.

[5] Appelhans, B. M. and Luecken, L. J. [2006]. Heart rate variability as an index of regulated emotional responding., *Review of general psychology* **10**(3): 229.

[6] Atkinson, R. C. and Shiffrin, R. M. [1968a]. Human memory: A proposed system and its control processes, *Psychology of learning and motivation* **2**: 89–195.

[7] Atkinson, R. C. and Shiffrin, R. M. [1968b]. Human memory: A proposed system and its control processes, *Psychology of learning and motivation* **2**: 89–195.

[8] Atkinson, R. C. and Shiffrin, R. M. [1971]. *The control processes of short-term memory*, Stanford University Stanford.

[9] Ausubel, D. P. and Youssef, M. [1965]. The effect of spaced repetition on meaningful retention, *The Journal of General Psychology* **73**(1): 147–150.

[10] Azcarate, A., Hageloh, F., Van de Sande, K. and Valenti, R. [2005]. Automatic facial emotion recognition, *Universiteit van Amsterdam* pp. 1–6.

[11] Baddeley, A. D. [1997]. *Human memory: Theory and practice*, Psychology Press.

[12] Baddeley, A. D. and Hitch, G. [1974]. Working memory, *Psychology of learning and motivation* **8**: 47–89.

[13] Baddeley, A., Hatter, J., Scott, D. and Snashall, A. [1970]. Memory and time of day, *The Quarterly Journal of Experimental Psychology* **22**(4): 605–609.

[14] Bahrainian, S. A. and Crestani, F. [2016]. Cued retrieval of personal memories of social interactions, *Proceedings of the first Workshop on Lifelogging Tools and Applications*, ACM, pp. 3–12.

[15] Bahrick, H. P., Bahrick, P. O. and Wittlinger, R. P. [1975]. Fifty years of memory for names and faces: A cross-sectional approach., *Journal of experimental psychology: General* **104**(1): 54.

[16] Bamidis, P. D., Papadelis, C., Kourtidou-Papadeli, C., Pappas, C. and B. Vivas, A. [2004]. Affective computing in the era of contemporary neurophysiology and health informatics, *Interacting with Computers* **16**(4): 715–721.

[17] Bangor, A., Kortum, P. T. and Miller, J. T. [2008]. An empirical evaluation of the system usability scale, *Intl. Journal of Human–Computer Interaction* **24**(6): 574–594.

[18] Barnier, A. J., Conway, M. A., Mayoh, L., Speyer, J., Avizmil, O. and Harris, C. B. [2007]. Directed forgetting of recently recalled autobiographical memories., *Journal of Experimental Psychology: General* **136**(2): 301.

[19] Barsalou, L. W. [1988]. The content and organization of autobiographical memories, *Remembering reconsidered: Ecological and traditional approaches to the study of memory* pp. 193–243.

[20] Bartlett, F. C. and Burt, C. [1933]. Remembering: A study in experimental and social psychology, *British Journal of Educational Psychology* **3**(2): 187–192.

[21] Berger, T. W., Hampson, R. E., Song, D., Goonawardena, A., Marmarelis, V. Z. and Deadwyler, S. A. [2011]. A cortical neural prosthesis for restoring and enhancing memory, *Journal of neural engineering* **8**(4): 046017.

[22] Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D. and Craven, P. L. [2007]. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks, *Aviation, space, and environmental medicine* **78**(5): B231–B244.

[23] Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B. and Wood, K. [2007]. The use of a wearable camera, sensecam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report, *Neuropsychological Rehabilitation* **17**(4-5): 582–601.

[24] Bexheti, A., Langheinrich, M. and Clinch, S. [2016]. Secure personal memory-sharing with co-located people and places, *Proceedings of the 6th International Conference on the Internet of Things*, ACM, pp. 73–81.

[25] Blei, D. M., Ng, A. Y., Jordan, M. I. and Lafferty, J. [2003]. Latent dirichlet allocation., *Journal of Machine Learning Research* **3**(4/5): 993 – 1022.

[26] Bliss, T. V., Collingridge, G. L. et al. [1993]. A synaptic model of memory: long-term potentiation in the hippocampus, *Nature* **361**(6407): 31–39.

[27] Böhmer, M., Hecht, B., Schöning, J., Krüger, A. and Bauer, G. [2011]. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage, *Proceedings of the 13th international conference on Human computer interaction with mobile devices and services*, ACM, pp. 47–56.

[28] Bohn, J., Coroamă, V., Langheinrich, M., Mattern, F. and Rohs, M. [2004]. Living in a world of smart everyday objects—social, economic, and ethical implications, *Human and Ecological Risk Assessment* **10**(5): 763–785.

[29] Brandimonte, M. A., Einstein, G. O. and McDaniel, M. A. [2014]. *Prospective memory: Theory and applications*, Psychology Press.

[30] Braun, V. and Clarke, V. [2006]. Using thematic analysis in psychology, *Qualitative research in psychology* **3**(2): 77–101.

[31] Bremner, J. D. and Narayan, M. [1998]. The effects of stress on memory and the hippocampus throughout the life cycle: Implications for childhood development and aging, *Development and psychopathology* **10**(04): 871–885.

[32] Brooke, J. et al. [1996]. Sus-a quick and dirty usability scale, *Usability evaluation in industry* **189**(194): 4–7.

[33] Burgess, P. W. and Shallice, T. [1997]. The relationship between prospective and retrospective memory: Neuropsychological evidence, *Cognitive models of memory* **5**: 249–256.

[34] Burt, D. B., Zembar, M. J. and Niederehe, G. [1995]. Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity., *Psychological bulletin* **117**(2): 285.

[35] Buschke, H. [1984]. Cued recall in amnesia, *Journal of Clinical and Experimental Neuropsychology* **6**(4): 433–440.

[36] Byrne, D., Lee, H., Jones, G. J. and Smeaton, A. F. [2008]. Guidelines for the presentation and visualisation of lifelog content.

[37] Cairns, P. and Cox, A. L. [2008]. *Research methods for human-computer interaction*, Vol. 12, Cambridge University Press Cambridge.

[38] Campbell, J. L., Quincy, C., Osserman, J. and Pedersen, O. K. [2013]. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement, *Sociological Methods & Research* **42**(3): 294–320.

[39] Caprani, N., Doherty, A. R., Lee, H., Smeaton, A. F., O'Connor, N. E. and Gurrin, C. [2010]. Designing a touch-screen sensecam browser to support an aging population, *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 4291–4296.

[40] Chen, Y. and Jones, G. J. [2010]. Augmenting human memory using personal lifelogs, *Proceedings of the 1st augmented human international conference*, ACM, p. 24.

[41] Clinch, S., Metzger, P. and Davies, N. [2014]. Lifelogging for'observer'view memories: an infrastructure approach, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM, pp. 1397–1404.

[42] Cohen, I., Sebe, N., Garg, A., Chen, L. S. and Huang, T. S. [2003]. Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and image understanding* **91**(1): 160–187.

[43] Cohen, M. D. and Bacdayan, P. [1994]. Organizational routines are stored as procedural memory: Evidence from a laboratory study, *Organization science* **5**(4): 554–568.

[44] Conner, T. S. and Reid, K. A. [2012]. Effects of intensive mobile happiness reporting in daily life, *Social Psychological and Personality Science* **3**(3): 315–323.

[45] Consolvo, S., Harrison, B., Smith, I., Chen, M. Y., Everitt, K., Froehlich, J. and Landay, J. A. [2007]. Conducting in situ evaluations for and with ubiquitous computing technologies, *International Journal of Human-Computer Interaction* **22**(1-2): 103–118.

[46] Consolvo, S. and Walker, M. [2003]. Using the experience sampling method to evaluate ubicomp applications, *IEEE Pervasive Computing* **2**(2): 24–31.

[47] Conway, M. A. [2005]. Memory and the self, *Journal of memory and language* **53**(4): 594–628.

[48] Conway, M. A. [2009]. Episodic memories, *Neuropsychologia* **47**(11): 2305–2313.

[49] Cook, P., Ellis, C., Graf, M., Rein, G. and Smith, T. [1987]. Project nick: meetings augmentation and analysis, *ACM Transactions on Information Systems (TOIS)* **5**(2): 132–146.

[50] Cook, T. D., Campbell, D. T. and Day, A. [1979]. *Quasi-experimentation: Design & analysis issues for field settings*, Vol. 351, Houghton Mifflin Boston.

[51] Corteen, R. [1969]. Skin conductance changes and word recall, *British Journal of Psychology* **60**(1): 81–84.

[52] Coyle, J. T., Price, D. L. and DeLong, M. R. [1983]. Alzheimer's disease: a disorder of cortical cholinergic innervation, *Science* **219**(4589): 1184–1190.

[53] Craik, F. I. and Lockhart, R. S. [1972]. Levels of processing: A framework for memory research, *Journal of verbal learning and verbal behavior* **11**(6): 671–684.

[54] Csikszentmihalyi, M. and Larson, R. [2014]. Validity and reliability of the experience-sampling method, *Flow and the foundations of positive psychology*, Springer, pp. 35–54.

[55] Cuddy, L. J. and Jacoby, L. L. [1982]. When forgetting helps memory: An analysis of repetition effects, *Journal of Verbal Learning and Verbal Behavior* **21**(4): 451–467.

[56] Davies, N., Friday, A., Clinch, S., Sas, C., Langheinrich, M., Ward, G. and Schmidt, A. [2015]. Security and privacy implications of pervasive memory augmentation, *IEEE Pervasive Computing* **14**(1): 44–53.

[57] De Guzman, E. S., Sharmin, M. and Bailey, B. P. [2007]. Should i call now? understanding what context is considered when deciding whether to initiate remote communication via mobile devices, *Proceedings of Graphics interface 2007*, ACM, pp. 143–150.

[58] Diener, E. and Tay, L. [2014]. Review of the day reconstruction method (drm), *Social Indicators Research* **116**(1): 255–267.

[59] Dingler, T., Bexheti, A., Niforatos, E. and Alt, F. [2015]. Workshop on mobile cognition: Using mobile devices to enhance human cognition, *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ACM, pp. 970–973.

[60] Dingler, T., El Agroudy, P., Le, H. V., Schmidt, A., Niforatos, E., Bexheti, A. and Langheinrich, M. [2016]. Multimedia memory cues for augmenting human memory, *IEEE MultiMedia* **23**(2): 4–11.

[61] Dingler, T., Kunze, K., Niforatos, E., Gurrin, C., Giannopolos, I., Dengel, A. and Kise, K. [2016]. Wahm 2016: 3rd workshop on ubiquitous technologies for augmenting the human mind, *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ACM, pp. 1010–1013.

[62] Dix, A. [2009]. Human-computer interaction, *Encyclopedia of database systems*, Springer, pp. 1327–1331.

[63] Dobbelstein, D., Rukzio, E. and Herrdum, S. [2017]. Demonstration of inscent: a wearable olfactory display as an amplification for mobile notifications, *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ACM, pp. 229–232.

[64] Dobbins, C., Merabti, M., Fergus, P. and Llewellyn-Jones, D. [2014]. The big data obstacle of lifelogging, *Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on*, IEEE, pp. 922–926.

[65] Dodge, M. and Kitchin, R. [2007]. 'outlines of a world coming into existence': pervasive computing and the ethics of forgetting, *Environment and planning B: planning and design* **34**(3): 431–445.

[66] Doherty, A. R., Pauly-Takacs, K., Caprani, N., Gurrin, C., Moulin, C. J., O'Connor, N. E. and Smeaton, A. F. [2012]. Experiences of aiding autobiographical memory using the sensecam, *Human–Computer Interaction* **27**(1-2): 151–174.

[67] Ehlen, P., Purver, M., Niekrasz, J., Lee, K. and Peters, S. [2008]. Meeting adjourned: off-line learning interfaces for automatic meeting understanding, *Proceedings of the 13th international conference on Intelligent user interfaces*, ACM, pp. 276–284.

[68] Einstein, G. O. and McDaniel, M. A. [1990]. Normal aging and prospective memory., *Journal of Experimental Psychology: Learning, Memory, and Cognition* **16**(4): 717.

[69] Ekman, P. [1993]. Facial expression and emotion., *American psychologist* **48**(4): 384.

[70] Elsden, C., Kirk, D. S. and Durrant, A. C. [2016]. A quantified past: Toward design for remembering with personal informatics, *Human–Computer Interaction* **31**(6): 518–557.

[71] Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., Kim, J. S., Heo, S., Alves, H., White, S. M. et al. [2011]. Exercise training increases size of hippocampus and improves memory, *Proceedings of the National Academy of Sciences* **108**(7): 3017–3022.

[72] Ericsson, K. A. and Simon, H. A. [1993]. *Protocol analysis*, MIT press Cambridge, MA.

[73] Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B. and Nguyen-Thien, N. [2010]. Emotion on the road – necessity, acceptance, and feasibility of affective computing in the car, *Advances in human-computer interaction* **2010**.

[74] Ezzyat, Y., Kragel, J. E., Burke, J. F., Levy, D. F., Lyalenko, A., Wanda, P., O'Sullivan, L., Hurley, K. B., Busygin, S., Pedisich, I. et al. [2017]. Direct brain stimulation modulates encoding states and memory performance in humans, *Current Biology* **27**(9): 1251–1258.

[75] Fedosov, A., Niforatos, E., Alt, F. and Elhart, I. [2015]. Supporting interactivity on a ski lift, *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ACM, pp. 767–770.

[76] Festinger, L. [1962]. *A theory of cognitive dissonance*, Vol. 2, Stanford university press.

[77] Field, A. [2013]. *Discovering statistics using IBM SPSS statistics*, Sage.

[78] Fisher, R. P., Geiselman, R. E. and Amador, M. [1989]. Field test of the cognitive interview: enhancing the recollection of actual victims and witnesses of crime., *Journal of Applied Psychology* **74**(5): 722.

[79] Fortin, N. J., Agster, K. L. and Eichenbaum, H. B. [2002]. Critical role of the hippocampus in memory for sequences of events, *Nature neuroscience* **5**(5): 458–462.

[80] Fowles, D. C. [1980]. The three arousal model: Implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy, *Psychophysiology* **17**(2): 87–104.

[81] Gaggioli, A., Ferscha, A., Riva, G., Dunne, S. and Viaud-Delmon, I. [2016]. Human computer confluence transforming human experience through symbiotic technologies.

[82] Gais, S., Lucas, B. and Born, J. [2006]. Sleep after learning aids memory recall, *Learning & Memory* **13**(3): 259–262.

[83] Gemmell, J., Bell, G. and Lueder, R. [2006]. Mylifebits: a personal database for everything, *Communications of the ACM* **49**(1): 88–95.

[84] Gemmell, J., Bell, G., Lueder, R., Drucker, S. and Wong, C. [2002]. Mylifebits: fulfilling the memex vision, *Proceedings of the tenth ACM international conference on Multimedia*, ACM, pp. 235–238.

[85] Gouveia, R. and Karapanos, E. [2013]. Footprint tracker: supporting diary studies with lifelogging, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2921–2930.

[86] Graf, P. and Schacter, D. L. [1985]. Implicit and explicit memory for new associations in normal and amnesic subjects., *Journal of Experimental Psychology: Learning, memory, and cognition* **11**(3): 501.

[87] Gray, J. A. [1990]. Brain systems that mediate both emotion and cognition, *Cognition & emotion* **4**(3): 269–288.

[88] Greco, A., Valenza, G. and Scilingo, E. P. [n.d.]. Advances in electrodermal activity processing with applications for mental health.

[89] Green, M. C., Hilken, J., Friedman, H., Grossman, K., Gasiewskj, J., Adler, R. and Sabini, J. [2005]. Communication via instant messenger: Short-and long-term effects, *Journal of Applied Social Psychology* **35**(3): 445–462.

[90] Grober, E., Buschke, H., Crystal, H., Bang, S. and Dresner, R. [1988]. Screening for dementia by memory testing, *Neurology* **38**(6): 900–900.

[91] Gruebler, A. and Suzuki, K. [2010]. Measurement of distal emg signals using a wearable device for reading facial expressions, *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, IEEE, pp. 4594–4597.

[92] Guan, Z., Lee, S., Cuddihy, E. and Ramey, J. [2006]. The validity of the stimulated retrospective think-aloud method as measured by eye tracking, *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, pp. 1253–1262.

[93] Gurrin, C., Smeaton, A. F. and Doherty, A. R. [2014]. Lifelogging: Personal big data, *Foundations and trends in information retrieval* **8**(1): 1–125.

[94] Gurrin, C., Smeaton, A. F., Qiu, Z. and Doherty, A. [2013]. Exploring the technical challenges of large-scale lifelogging, *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference*, ACM, pp. 68–75.

[95] Haapalainen, E., Kim, S., Forlizzi, J. F. and Dey, A. K. [2010]. Psychophysiological measures for assessing cognitive load, *Proceedings of the 12th ACM international conference on Ubiquitous computing*, ACM, pp. 301–310.

[96] Hamann, S. [2001]. Cognitive and neural mechanisms of emotional memory, *Trends in cognitive sciences* **5**(9): 394–400.

[97] Harris, H. and Nass, C. [2011]. Emotion regulation for frustrating driving contexts, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 749–752.

[98] Harris, J. A., Miniussi, C., Harris, I. M. and Diamond, M. E. [2002]. Transient storage of a tactile memory trace in primary somatosensory cortex, *The Journal of Neuroscience* **22**(19): 8720–8725.

[99] Hart, S. G. [2006]. Nasa-task load index (nasa-tlx); 20 years later, *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50, Sage Publications Sage CA: Los Angeles, CA, pp. 904–908.

[100] Hart, S. G. and Staveland, L. E. [1988]. Development of nasa-tlx (task load index): Results of empirical and theoretical research, *Advances in psychology* **52**: 139–183.

[101] Hartson, R. and Pyla, P. S. [2012]. *The UX Book: Process and guidelines for ensuring a quality user experience*, Elsevier.

[102] Harvey, M., Langheinrich, M. and Ward, G. [2016]. Remembering through lifelogging: A survey of human memory augmentation, *Pervasive and Mobile Computing* **27**: 14–26.

[103] Hassenzahl, M. [2008]. User experience (ux): towards an experiential perspective on product quality, *Proceedings of the 20th Conference on l'Interaction Homme-Machine*, ACM, pp. 11–15.

[104] Hassenzahl, M. [2013]. User experience and experience design, *The Encyclopedia of Human-Computer Interaction*, .

[105] Hassenzahl, M. and Tractinsky, N. [2006]. User experience-a research agenda, *Behaviour & information technology* **25**(2): 91–97.

[106] Hearst, M. A. [1997]. Texttiling: Segmenting text into multi-paragraph subtopic passages, *Comput. Linguist.* **23**(1): 33–64.
URL: *http://dl.acm.org/citation.cfm?id=972684.972687*

[107] Heindel, W. C., Salmon, D. P. and Butters, N. [1990]. Pictorial priming and cued recall in alzheimer's and huntington's disease, *Brain and Cognition* **13**(2): 282–295.

[108] Henkel, L. A. [2014]. Point-and-shoot memories the influence of taking photos on memory for a museum tour, *Psychological science* **25**(2): 396–402.

[109] Herz, R. S. [2004]. A naturalistic analysis of autobiographical memories triggered by olfactory visual and auditory stimuli, *Chemical Senses* **29**(3): 217–224.

[110] Hirschberg, A. W., Xu, H., Scholien, K., Berger, T. W., Song, D. and Meng, E. [2017]. Development of an anatomically conformal parylene neural probe array for multi-region hippocampal recordings, *Micro Electro Mechanical Systems (MEMS), 2017 IEEE 30th International Conference on*, IEEE, pp. 129–132.

[111] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N. and Wood, K. [2006]. Sensecam: A retrospective memory aid, *UbiComp 2006: Ubiquitous Computing* pp. 177–193.

[112] Hofman, M. A. [2014]. Evolution of the human brain: when bigger is better, *Frontiers in neuroanatomy* **8**.

[113] Holmes, E. A., Brewin, C. R. and Hennessy, R. G. [2004]. Trauma films, information processing, and intrusive memory development., *Journal of Experimental Psychology: General* **133**(1): 3.

[114] Hopfgartner, F., Yang, Y., Zhou, L. M. and Gurrin, C. [2013]. User interaction templates for the design of lifelogging systems, *Semantic Models for Adaptive Interactive Systems*, Springer, pp. 187–204.

[115] Hunter, S., Maes, P., Scott, S. and Kaufman, H. [2011]. Memtable: an integrated system for capture and recall of shared histories in group workspaces, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 3305–3314.

[116] Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J.-H. and Dey, A. K. [2012]. Factors influencing quality of experience of commonly used mobile applications, *IEEE Communications Magazine* **50**(4).

[117] Isaacs, E., Konrad, A., Walendowski, A., Lennig, T., Hollis, V. and Whittaker, S. [2013]. Echoes from the past: how technology mediated reflection improves well-being, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 1071–1080.

[118] Jaeggi, S. M., Buschkuehl, M., Jonides, J. and Shah, P. [2011]. Short-and long-term benefits of cognitive training, *Proceedings of the National Academy of Sciences* **108**(25): 10081–10086.

[119] Jaimes, A., Omura, K., Nagamine, T. and Hirata, K. [2004]. Memory cues for meeting video retrieval, *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, ACM, pp. 74–85.

[120] Jin, B. and Peña, J. F. [2010]. Mobile communication in romantic relationships: Mobile phone use, relational uncertainty, love, commitment, and attachment styles, *Communication Reports* **23**(1): 39–51.

[121] Joslyn, S. L. and Oakes, M. A. [2005]. Directed forgetting of autobiographical events, *Memory & cognition* **33**(4): 577–587.

[122] Josselyn, S. A. [2010]. Continuing the search for the engram: examining the mechanism of fear memories, *J Psychiatry Neurosci* **35**(4): 221–8.

[123] Kahana, M. J., Howard, M. W. and Polyn, S. M. [2008]. Associative retrieval processes in episodic memory.

[124] Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. and Stone, A. A. [2004]. A survey method for characterizing daily life experience: The day reconstruction method, *Science* **306**(5702): 1776–1780.

[125] Kalnikaite, V., Sellen, A., Whittaker, S. and Kirk, D. [2010]. Now let me see where i was: understanding how lifelogs mediate memory, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2045–2054.

[126] Kalnikaitė, V. and Whittaker, S. [2011]. A saunter down memory lane: Digital reflection on personal mementos, *International Journal of Human-Computer Studies* **69**(5): 298–310.

[127] Kalnikaite, V. and Whittaker, S. [2012]. Synergetic recollection: how to design lifelogging tools that help locate the right information, *Human-Computer Interaction: The Agency Perspective* pp. 329–348.

[128] Karapanos, E., Barreto, M., Nisi, V. and Niforatos, E. [2012]. Does locality make a difference? assessing the effectiveness of location-aware narratives, *Interacting with Computers* **24**(4): 273–279.

[129] Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J.-B. [2009]. User experience over time: an initial framework, *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, pp. 729–738.

[130] Kenny, D. A. and Acitelli, L. K. [2001]. Accuracy and bias in the perception of the partner in a close relationship, *Journal of personality and social psychology* **80**(3): 439–448.

[131] Kern, D. and Schmidt, A. [2009]. Design space for driver-based automotive user interfaces, *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ACM, pp. 3–10.

[132] Killingsworth, M. A. and Gilbert, D. T. [2010]. A wandering mind is an unhappy mind, *Science* **330**(6006): 932–932.

[133] Kleinsmith, L. J., Kaplan, S. and Trate, R. D. [1963]. The relationship of arousal to short-and long-term verbal recall., *Canadian Journal of Psychology/Revue canadienne de psychologie* **17**(4): 393.

[134] Klimesch, W. [1999]. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis, *Brain research reviews* **29**(2): 169–195.

[135] Klingberg, T. [2010]. Training and plasticity of working memory, *Trends in cognitive sciences* **14**(7): 317–324.

[136] Krippendorff, K. and Craggs, R. [2016]. The reliability of multi-valued coding of data, *Communication Methods and Measures* **10**(4): 181–198.

[137] Krumm, J. [2012]. How people use their vehicles: Statistics from the 2009 national household travel survey, *Technical report*, SAE Technical Paper.

[138] LaBar, K. S. and Cabeza, R. [2006]. Cognitive neuroscience of emotional memory, *Nature Reviews Neuroscience* **7**(1): 54–64.

[139] Lajunen, T. and Parker, D. [2001]. Are aggressive people aggressive drivers? a study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving, *Accident Analysis & Prevention* **33**(2): 243–255.

[140] Langheinrich, M., Schmidt, A., Davies, N. and José, R. [2013]. A practical framework for ethics: the pd-net approach to supporting ethics compliance in public display studies, *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, ACM, pp. 139–143.

[141] Larson, R. [1989]. Is feeling "in control" related to happiness in daily life?, *Psychological Reports* **64**(3): 775–784.

[142] Larson, R. and Csikszentmihalyi, M. [1983]. The experience sampling method., *New Directions for Methodology of Social & Behavioral Science* .

[143] Lazar, J., Feng, J. H. and Hochheiser, H. [2017]. *Research methods in human-computer interaction*, Morgan Kaufmann.

[144] Lee, M. L. and Dey, A. K. [2007]. Providing good memory cues for people with episodic memory impairment, *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, ACM, pp. 131–138.

[145] Lee, M. L. and Dey, A. K. [2008]. Lifelogging memory appliance for people with episodic memory impairment, *Proceedings of the 10th international conference on Ubiquitous computing*, ACM, pp. 44–53.

[146] Levine, B., Svoboda, E., Hay, J. F., Winocur, G. and Moscovitch, M. [2002]. Aging and autobiographical memory: dissociating episodic from semantic retrieval., *Psychology and aging* **17**(4): 677.

[147] Lindley, S. E. [2012]. Before i forget: From personal memory to family history, *Human–Computer Interaction* **27**(1-2): 13–36.

[148] Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K. and Tonegawa, S. [2012]. Optogenetic stimulation of a hippocampal engram activates fear memory recall, *Nature* **484**(7394): 381–385.

[149] Lynn, M. [1992]. The psychology of unavailability: Explaining scarcity and cost effects on value, *Basic and Applied Social Psychology* **13**(1): 3–7.

[150] Machajdik, J., Hanbury, A., Garz, A. and Sablatnig, R. [2011]. Affective computing for wearable diary and lifelogging systems: An overview, *Machine Vision-Research for High Quality Processes and Products-35th Workshop of the Austrian Association for Pattern Recognition. Austrian Computer Society*.

[151] MacKerron, G. and Mourato, S. [2010]. Lse's mappiness project may help us track the national mood: but how much should we consider happiness in deciding public policy?, *British Politics and Policy at LSE* .

[152] Maltzman, I., Kantor, W. and Langdon, B. [1966]. Immediate and delayed retention, arousal, and the orienting and defensive reflexes, *Psychonomic Science* **6**(10): 445–446.

[153] Marchiori, E., Niforatos, E. and Preto, L. [2017]. Measuring the media effects of a tourism-related virtual reality experience using biophysical data, *Information and Communication Technologies in Tourism 2017*, Springer, pp. 203–215.

[154] Marchiori, E., Niforatos, E. and Preto, L. [n.d.]. Analysis of users' heart rate data and self-reported perceptions to understand effective virtual reality characteristics, *Information Technology & Tourism* pp. 1–23.

[155] Martin, A. and Chao, L. L. [2001]. Semantic memory and the brain: structure and processes, *Current opinion in neurobiology* **11**(2): 194–201.

[156] Matzen, L. E., Trumbo, M. C., Leach, R. C. and Leshikar, E. D. [2015]. Effects of non-invasive brain stimulation on associative memory, *Brain research* **1624**: 286–296.

[157] McCarthy, J. and Wright, P. [2004]. Technology as experience, *interactions* **11**(5): 42–43.

[158] McDaniel, M. A. and Masson, M. E. [1985]. Altering memory representations through retrieval., *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11**(2): 371.

[159] Meeter, M. and Murre, J. M. [2004]. Consolidation of long-term memory: evidence and alternatives., *Psychological Bulletin* **130**(6): 843.

[160] Miller, G. A. and Chapman, J. P. [2001]. Misunderstanding analysis of covariance., *Journal of abnormal psychology* **110**(1): 40.

[161] Mitchell, T. R., Thompson, L., Peterson, E. and Cronk, R. [1997]. Temporal adjustments in the evaluation of events: The "rosy view", *Journal of Experimental Social Psychology* **33**(4): 421–448.

[162] Murdock Jr, B. B. [1962]. The serial position effect of free recall., *Journal of experimental psychology* **64**(5): 482.

[163] Newgard, C. D. and Haukoos, J. S. [2007]. Advanced statistics: missing data in clinical research—part 2: multiple imputation, *Academic Emergency Medicine* **14**(7): 669–678.

[164] Niforatos, E., Cinel, C., Mack, C. C., Langheinrich, M. and Ward, G. [2017]. Can less be more?: Contrasting limited, unlimited, and automatic picture capture for augmenting memory recall, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**(2): 21.

[165] Niforatos, E. and Karapanos, E. [2015]. Emosnaps: a mobile application for emotion recall from facial expressions, *Personal and Ubiquitous Computing* **19**(2): 425–444.

[166] Niforatos, E., Karapanos, E., Langheinrich, M., Wurhofer, D., Krischkowsky, A., Obrist, M. and Tscheligi, M. [2015]. emotion: retrospective in-car user experience evaluation, *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ACM, pp. 118–123.

[167] Niforatos, E., Langheinrich, M. and Bexheti, A. [2014]. My good old kodak: understanding the impact of having only 24 pictures to take, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM, pp. 1355–1360.

[168] Niforatos, E., Lim, V., Vuerich, C., Langheinrich, M. and Bexheti, A. [2015]. Pulsecam: Biophysically driven life logging, *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, ACM, pp. 1002–1009.

[169] Niforatos, E., Matías, L., Bexheti, A. and Langheinrich, M. [2018]. Augmenting memory recall in work meetings: Establishing a antifiable baseline, *Proceedings of the 9th Augmented Human International Conference*, ACM.

[170] Niforatos, E., Vourvopoulos, A. and Langheinrich, M. [2017]. Amplifying human cognition: bridging the cognitive gap between human and machine, *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ACM, pp. 673–680.

[171] Nightingale, S., Wade, K., Watson, D., Mills, A., Zajac, R. and Garry, M. [2016]. Taking photos does not impair memory, *Proc. of the 6th Intl. Conference on Memory*, Budapest, Hungary, pp. 648–657.

[172] Ohl, F. and Fuchs, E. [1999]. Differential effects of chronic stress on memory processes in the tree shrew, *Cognitive Brain Research* **7**(3): 379–387.

[173] Ortony, A. and Turner, T. J. [1990]. What's basic about basic emotions?, *Psychological review* **97**(3): 315.

[174] Oulasvirta, A., Rattenbury, T., Ma, L. and Raita, E. [2012]. Habits make smartphone use more pervasive, *Personal and Ubiquitous Computing* **16**(1): 105–114.

[175] Palomba, D., Angrilli, A. and Mini, A. [1997]. Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli, *International journal of psychophysiology* **27**(1): 55–67.

[176] Pantic, M. and Rothkrantz, L. J. M. [2000]. Automatic analysis of facial expressions: The state of the art, *IEEE Transactions on pattern analysis and machine intelligence* **22**(12): 1424–1445.

[177] Pavlov, I. P. and Anrep, G. V. [2003]. *Conditioned reflexes*, Courier Corporation.

[178] Petrelli, D., Villar, N., Kalnikaite, V., Dib, L. and Whittaker, S. [2010]. Fm radio: family interplay with sonic mementos, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2371–2380.

[179] Petrelli, D. and Whittaker, S. [2010]. Family memories in the home: contrasting physical and digital mementos, *Personal and Ubiquitous Computing* **14**(2): 153–169.

[180] Petrelli, D., Whittaker, S. and Brockmeier, J. [2008]. Autotopography: what can physical mementos tell us about digital memories?, *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, pp. 53–62.

[181] Phelps, E. A. [2004]. Human emotion and memory: interactions of the amygdala and hippocampal complex, *Current opinion in neurobiology* **14**(2): 198–202.

[182] Pielot, M., Dingler, T., Pedro, J. S. and Oliver, N. [2015]. When attention is not scarce-detecting boredom from mobile phone usage, *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, ACM, pp. 825–836.

[183] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. and Podsakoff, N. P. [2003]. Common method biases in behavioral research: A critical review of

the literature and recommended remedies., *Journal of applied psychology* **88**(5): 879.

[184] Qiu, Z., Gurrin, C., Doherty, A. R. and Smeaton, A. F. [2012]. A real-time life experience logging tool, *International Conference on Multimedia Modeling*, Springer, pp. 636–638.

[185] Quillan, M. R. [1966]. Semantic memory, *Technical report*, BOLT BERANEK AND NEWMAN INC CAMBRIDGE MA.

[186] Racsmány, M. and Conway, M. A. [2006]. Episodic inhibition., *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**(1): 44.

[187] Rapley, T. J. [2001]. The art (fulness) of open-ended interviewing: some considerations on analysing interviews, *Qualitative research* **1**(3): 303–323.

[188] Reeves, A. and Sperling, G. [1986]. Attention gating in short-term visual memory., *Psychological review* **93**(2): 180.

[189] Robinson, M. D. and Clore, G. L. [2002]. Belief and feeling: evidence for an accessibility model of emotional self-report., *Psychological bulletin* **128**(6): 934.

[190] Rowe, D. W., Sibert, J. and Irwin, D. [1998]. Heart rate variability: Indicator of user state as an aid to human-computer interaction, *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co., pp. 480–487.

[191] Russel, J. A., Weiss, A. and Mendelsohn, G. A. [1989]. Affect grid: A single-item scale of pleasure and arousal, *Journal of Personality and Social Psychology* **57**(3): 493–502.

[192] Russell, J. A. [1994]. Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies., *Psychological bulletin* **115**(1): 102.

[193] Ryan, L., Cox, C., Hayes, S. M. and Nadel, L. [2008]. Hippocampal activation during episodic and semantic memory retrieval: Comparing category production and category cued recall, *Neuropsychologia* **46**(8): 2109–2121.

[194] Ryan, R. M., Bernstein, J. H. and Brown, K. W. [2010]. Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms, *Journal of social and clinical psychology* **29**(1): 95–122.

[195] Rystrom, D. S. and Benson, E. D. [1989]. Investor psychology and the day-of-the-week effect, *Financial Analysts Journal* **45**(5): 75–78.

[196] Ryu, K. and Myung, R. [2005]. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic, *International Journal of Industrial Ergonomics* **35**(11): 991–1009.

[197] Sanches, P., Höök, K., Vaara, E., Weymann, C., Bylund, M., Ferreira, P., Peira, N. and Sjölinder, M. [2010]. Mind the body!: designing a mobile stress management application encouraging personal reflection, *Proceedings of the 8th ACM conference on designing interactive systems*, ACM, pp. 47–56.

[198] Sanchez-Vives, M. V. and Slater, M. [2005]. From presence to consciousness through virtual reality, *Nature Reviews Neuroscience* **6**(4): 332.

[199] Sas, C., Challioner, S., Clarke, C., Wilson, R., Coman, A., Clinch, S., Harding, M. and Davies, N. [2015]. Self-defining memory cues: creative expression and emotional meaning, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 2013–2018.

[200] Sas, C., Fratczak, T., Rees, M., Gellersen, H., Kalnikaite, V., Coman, A. and Höök, K. [2013]. Affectcam: arousal-augmented sensecam for richer recall of episodic memories, *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 1041–1046.

[201] Sas, C. and Whittaker, S. [2013]. Design for forgetting: disposing of digital possessions after a breakup, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 1823–1832.

[202] Schacter, D. L. [1999]. The seven sins of memory: Insights from psychology and cognitive neuroscience., *American psychologist* **54**(3): 182.

[203] Schaefer, A. and Philippot, P. [2005]. Selective effects of emotion on the phenomenal characteristics of autobiographical memories, *Memory* **13**(2): 148–160.

[204] Scherer, R., Friedrich, E. C., Allison, B., Pröll, M., Chung, M., Cheung, W., Rao, R. P. and Neuper, C. [2011]. Non-invasive brain-computer interfaces: Enhanced gaming and robotic control, *International Work-Conference on Artificial Neural Networks*, Springer, pp. 362–369.

[205] Schmidt, A. [2017]. Technologies to amplify the mind, *Computer* **50**(10): 102–106.

[206] Schmidt, A., Dey, A. K., Kun, A. L. and Spiessl, W. [2010]. Automotive user interfaces: human computer interaction in the car, *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 3177–3180.

[207] Schoenemann, P. T. [2006]. Evolution of the size and functional areas of the human brain, *Annu. Rev. Anthropol.* **35**: 379–406.

[208] Scollon, C. N., Prieto, C.-K. and Diener, E. [2009]. Experience sampling: promises and pitfalls, strength and weaknesses, *Assessing well-being*, Springer, pp. 157–180.

[209] Sellen, A. J., Fogg, A., Aitken, M., Hodges, S., Rother, C. and Wood, K. [2007]. Do life-logging technologies support memory for the past?: an experimental study using sensecam, *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp. 81–90.

[210] Sellen, A. J. and Whittaker, S. [2010]. Beyond total capture: a constructive critique of lifelogging, *Communications of the ACM* **53**(5): 70–77.

[211] Soares, J. S. and Storm, B. C. [2017]. Forget in a flash: A further investigation of the photo-taking-impairment effect, *Journal of Applied Research in Memory and Cognition* .

[212] Sparrow, B., Liu, J. and Wegner, D. M. [2011]. Google effects on memory: Cognitive consequences of having information at our fingertips, *science* **333**(6043): 776–778.

[213] Ståhl, A., Höök, K., Svensson, M., Taylor, A. S. and Combetto, M. [2009]. Experiencing the affective diary, *Personal and Ubiquitous Computing* **13**(5): 365–378.

[214] Stemmler, G., Spielberger, C. and Potegal, M. [2010]. International handbook of anger: Constituent and concomitant biological, psychological, and social processes.

[215] Stephens, A. N. and Groeger, J. A. [2011]. Anger-congruent behaviour transfers across driving situations, *Cognition & emotion* **25**(8): 1423–1438.

[216] Stone, A. A., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A. and Kahneman, D. [2006]. A population approach to the study of emotion: diurnal rhythms of a working day examined with the day reconstruction method., *Emotion* **6**(1): 139.

[217] Stone, A. A., Smyth, J. M., Pickering, T. and Schwartz, J. [1996]. Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns, *Journal of Applied Social Psychology* **26**(14): 1286–1305.

[218] Storm, B. C., Stone, S. M. and Benjamin, A. S. [2016]. Using the internet to access information inflates future use of the internet to access other information, *Memory* pp. 1–7.

[219] Suzuki, W. A. [2005]. Psychological science agenda| february 2005, *Psychological Science* .

[220] Taib, R., Tederry, J. and Itzstein, B. [2014]. Quantifying driver frustration to improve road safety, *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems*, ACM, pp. 1777–1782.

[221] Talarico, J. M., LaBar, K. S. and Rubin, D. C. [2004]. Emotional intensity predicts autobiographical memory experience, *Memory & cognition* **32**(7): 1118–1132.

[222] Teeters, A., El Kaliouby, R. and Picard, R. [2006]. Self-cam: feedback from what would be your social partner, *ACM SIGGRAPH 2006 Research posters*, ACM, p. 138.

[223] Templeton, G. F. et al. [2011]. A two-step approach for transforming continuous variables to normal: implications and recommendations for is research., *CAIS* **28**(1): 41–58.

[224] Thair, H., Holloway, A. L., Newport, R. and Smith, A. D. [2017]. Transcranial direct current stimulation (tdcs): A beginner's guide for design and implementation, *Frontiers in Neuroscience* **11**.

[225] Tulving, E. [1974]. Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible, *American Scientist* **62**(1): 74–82.

[226] Tulving, E. [1984]. Precis of elements of episodic memory, *Behavioral and Brain Sciences* **7**(2): 223–238.

[227] Tulving, E. [1986]. Episodic and semantic memory: Where should we go from here?, *Behavioral and Brain Sciences* **9**(3): 573–577.

[228] Tulving, E. and Markowitsch, H. J. [1998]. Episodic and declarative memory: role of the hippocampus, *Hippocampus* **8**(3): 198–204.

[229] Tulving, E. and Pearlstone, Z. [1966]. Availability versus accessibility of information in memory for words, *Journal of Verbal Learning and Verbal Behavior* **5**(4): 381–391.

[230] Tulving, E. and Thomson, D. M. [1973]. Encoding specificity and retrieval processes in episodic memory., *Psychological review* **80**(5): 352.

[231] Tulving, E. et al. [1972]. Episodic and semantic memory, *Organization of memory* **1**: 381–403.

[232] Tuokko, H., Vernon-Wilkinson, R., Weir, J. and Beattie, B. [1991]. Cued recall and early identification of dementia, *Journal of Clinical and Experimental Neuropsychology* **13**(6): 871–879.

[233] van Den Hoven, E., Sas, C. and Whittaker, S. [2012]. Introduction to this special issue on designing for personal memories: past, present, and future, *Human–Computer Interaction* **27**(1-2): 1–12.

[234] Verkasalo, H. [2009]. Contextual patterns in mobile service usage, *Personal and Ubiquitous Computing* **13**(5): 331–342.

[235] Vidal, J. J. [1973]. Toward direct brain-computer communication, *Annual review of Biophysics and Bioengineering* **2**(1): 157–180.

[236] Vourvopoulos, A., Ferreira, A. and i Badia, S. B. [2016]. Neurow: An immersive vr environment for motor-imagery training with the use of brain-computer interfaces and vibrotactile feedback., *PhyCS*, pp. 43–53.

[237] Vourvopoulos, A. and i Badia, S. B. [2016]. Usability and cost-effectiveness in brain-computer interaction: Is it user throughput or technology related?, *Proceedings of the 7th Augmented Human International Conference 2016*, ACM, p. 19.

[238] Walker, E. L. and Tarte, R. D. [1963]. Memory storage as a function of arousal and time with homogeneous and heterogeneous lists, *Journal of Verbal Learning and Verbal Behavior* **2**(1): 113–119.

[239] Ward, R. D. and Marsden, P. H. [2003]. Physiological responses to different web page designs, *International Journal of Human-Computer Studies* **59**(1): 199–212.

[240] Weinberg, L., Hasni, A., Shinohara, M. and Duarte, A. [2014]. A single bout of resistance exercise can enhance episodic memory performance, *Acta psychologica* **153**: 13–19.

[241] Wester, F. [2005]. K. krippendorff, content analysis: An introduction to its methodology , thousand oaks, ca: Sage, 2004 isbn 0-7619-1544-3.

[242] White, K. G. [2001]. Forgetting functions, *Learning & Behavior* **29**(3): 193–207.

[243] Whittaker, S., Kalnikaitė, V., Petrelli, D., Sellen, A., Villar, N., Bergman, O., Clough, P. and Brockmeier, J. [2012]. Socio-technical lifelogging: Deriving design principles for a future proof digital past, *Human–Computer Interaction* **27**(1-2): 37–62.

[244] Wickelgren, W. A. [1974]. Single-trace fragility theory of memory dynamics, *Memory & Cognition* **2**(4): 775–780.

[245] Wilson, G. F. [2002]. An analysis of mental workload in pilots during flight using multiple psychophysiological measures, *The International Journal of Aviation Psychology* **12**(1): 3–18.

[246] Witmer, B. G. and Singer, M. F. [1994]. Measuring presence in virtual environments, *Technical report*, ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA.

[247] Witmer, B. G. and Singer, M. J. [1998]. Measuring presence in virtual environments: A presence questionnaire, *Presence: Teleoperators and virtual environments* **7**(3): 225–240.

[248] Wixted, J. T. and Carpenter, S. K. [2007]. The wickelgren power law and the ebbinghaus savings function, *Psychological Science* **18**(2): 133–134.

[249] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. and Vaughan, T. M. [2002]. Brain–computer interfaces for communication and control, *Clinical neurophysiology* **113**(6): 767–791.

[250] Woodberry, E., Browne, G., Hodges, S., Watson, P., Kapur, N. and Woodberry, K. [2015]. The use of a wearable camera improves autobiographical memory in patients with alzheimer's disease, *Memory* **23**(3): 340–349.

[251] Wurhofer, D., Krischkowsky, A., Obrist, M., Karapanos, E., Niforatos, E. and Tscheligi, M. [2015]. Everyday commuting: prediction, actual experience and recall of anger and frustration in the car, *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ACM, pp. 233–240.

[252] Yang, Y., Lee, H. and Gurrin, C. [2013]. Visualizing lifelog data for different interaction platforms, *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM, pp. 1785–1790.

[253] Young, A. W., McWeeny, K. H., Hay, D. C. and Ellis, A. [1986]. Matching familiar and unfamiliar faces on identity and expression, *Psychological research* **48**(2): 63–68.

[254] Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M. and Atkins, M. S. [2012]. Workload assessment of surgeons: correlation between nasa tlx and blinks, *Surgical endoscopy* **26**(10): 2746–2750.

[255] Zhou, L. M. and Gurrin, C. [2012]. A survey on life logging data capturing, *SenseCam 2012* .