Footprint Tracker: Supporting Diary Studies with Lifelogging

Rúben Gouveia
Madeira Interactive Technologies Institute
Campus da Penteada
9020-105, Funchal Portugal
ruben.gouveia@m-iti.org

Evangelos Karapanos
Madeira Interactive Technologies Institute
Campus da Penteada
9020-105, Funchal Portugal
e.karapanos@m-iti.org

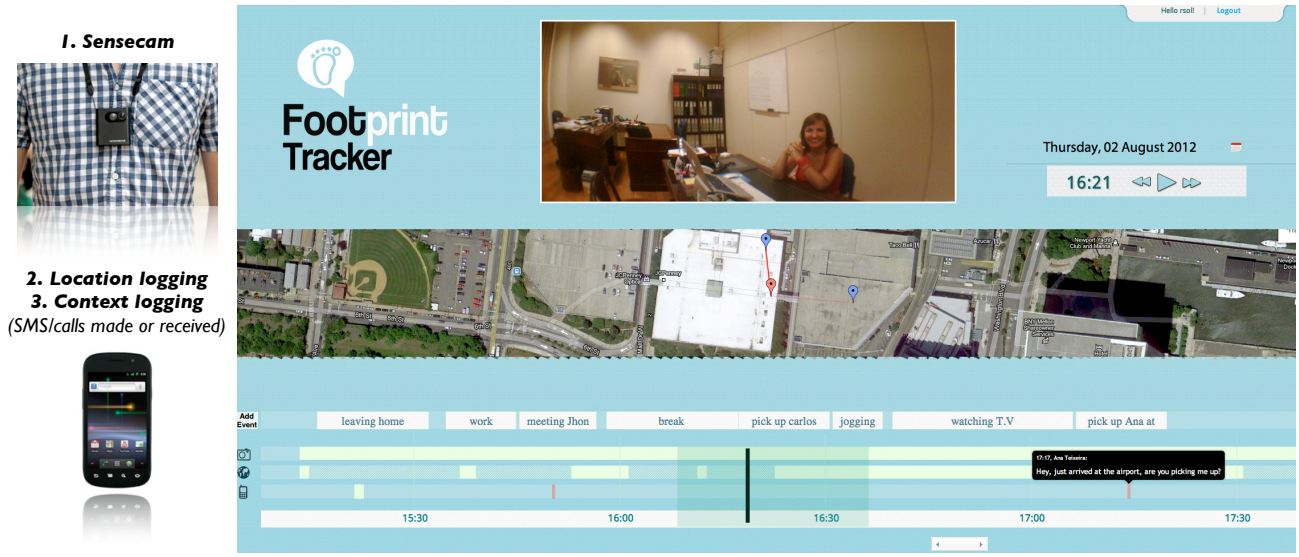

Figure 1. Footprint Tracker enables reviewing *visual*, *location*, *temporal* and *social context* cues, with the aim to recall and reflect upon daily activities and experiences in the context of diary studies.

ABSTRACT

As HCI shifts "to the wild", in-situ methods such as Diary Methods and the Experience Sampling Method are gaining momentum. However, researchers have acknowledged the *intrusiveness* and *lack of realism* in these methods and have proposed solutions, notably through lightweight and rich media capture. In this paper we explore the concept of lifelogging as an alternative solution to these two challenges. We describe *Footprint Tracker*, a tool that allows the review of lifelogs with the aim to support recall and reflection over daily activities and experiences. In a field trial, we study how four different types of cues, namely *visual*, *location*, *temporal* and *social context,* trigger memories of recent events and associated emotions. We conclude with a number of implications for the design of lifelogging systems that support recall and reflection upon recent events as well as ones lying further in our past.

Author Keywords

Diary methods; lifelogging; experience sampling.

ACM Classification Keywords

H5.2. User Interfaces: Evaluation/methodology.

INTRODUCTION

An increasing emphasis within HCI is paid to designing and evaluating technologies "in the wild" [6]. This has brought a shift towards methods for in-situ data collection that are appropriate for monitoring potentially *sparse* data, over *prolonged* periods of time and that remain relatively *unobtrusive* to participants' lives [8].

Interest in the Experience Sampling Method (ESM) [9] and Diary Methods (DM) [4], for example, has peaked over the last decade [17]. ESM and DM differ in the level of control the participant has in reporting. While participants in experience sampling studies respond to surveys that are introduced by the system either at random or

CHI 2013, April 27–May 2, 2013, Paris, France.

computationally estimated times, in diary studies, participants self-select when and what to report. Both methods, however, aim at sampling users' experiences and behaviors right at the moment when they occur. In doing so, they avoid retrospection and rationalization biases that are known in self-reporting methods [27].

While the experience sampling and diary methods are considered the gold standard of in-situ data collection [15], they also entail important drawbacks as they are *disruptive* to participants' daily activities and suffer from a lack of *realism* as the remote researcher does not have rich data about the situations on which participants report [8].

Alternative methods that address these two limitations have been proposed and adopted by the HCI community. For instance, Kahneman et al. [15] proposed the Day Reconstruction Method (DRM), a retrospective self-report protocol that aims at increasing participants' accuracy in reconstructing their experiences and whereabouts at the end of a studied day. It does so by imposing a chronological order in reconstruction, thus providing a temporal context to each recall. DRM has been found to provide a reasonably good approximation to experience sampling data [28], while it does not interrupt participants throughout the day and, being offline, it allows the elicitation of rich qualitative accounts [19].

Others have attempted to augment diary methods with rich media capture that are later used in cueing the recall of activities and experience. For instance, in photo-elicitation studies [6] participants are typically asked to capture a photo whenever they feel a need to capture something; during a retrospective interview, participants may use the captured photos to reminisce the encountered events and associated emotions. Carter and Mankoff [8] studied how different media such as photographs, audio recordings, location information and tangible artifacts cue the later recall of daily activities in the context of diary studies. They found photos to have a greater capacity than the other media in cueing the recall of detailed information, especially with respect to location and social context. Brandt et al. [5] studied how the capturing of small snippets of information in the field – in the form of text, photos and audio – assists the later recall in diary studies. They found that while the preferred medium of capturing varied per individual, photos were often instrumental in establishing the context of the sampled event while text was used to summarize the specific event and their motivations.

While these methods have provided significant advances in increasing realism in diary studies, they are still, to some extend, disruptive to participants' daily activities, as participants need to pay attention to and select which moments to sample.

In this paper we advocate for a methodological approach that is transparent in participants' daily lives and can be employed in long-term field studies. For that, we resort to the concept of *lifelogging*, the continuous capture of personal data such as photos from one's field-of-view, location, audio, biometric signals and others, with the aim of supporting the later recall and reflection over one's life events and experiences. The proposed approach offers two advantages over diary studies with rich media capture. First, it imposes no burden to participants as the system captures data continuously, rather than the user selecting when and what data to capture. Second, in doing so, it minimizes the risk of missing important moments, a common problem in diary studies.

This paper attempts three contributions. First, we describe *Footprint Tracker*, a system that was designed with the aim of supporting the review of lifelogs in the context of diary studies. Second, through a field trial of *Footprint Tracker* we study how four different types of cues, namely *visual*, *location*, *temporal* and *social context,* trigger memories of recent events and associated emotions. Third, we translate our findings to a set of implications for the design of lifelogging systems that support recall and reflection upon recent events as well as ones lying further in our past.

LIFELOGS AS MEMORY CUES

Memory was for long understood as a faithful account of past events, which can be reproduced when trying to remember details of the past. This idea was first challenged in Bartlett's [2] seminal work, which suggested that remembering is an act of *reconstruction* that can never produce the exact past event, but instead, every attempt to recall results in a new, often altered representation of the event. Tulving [32] proposed that recall happens through a process he called *synergistic ecphory*, a "largely preconscious process in which retrieval cues are brought in contact with stored information causing parts of that stored information to be reactivated". Tulving's model challenged a long-tenable idea, that different memory traces may have different strength. Instead, this model suggested that any given trace has many different strengths. Thus, different types of cues may have a different ability to cause different parts of a memory to be reactivated.

Importantly, recalling one's emotions and experiences associated with a given event follows the very same principles of episodic recall. Robinson and Clore [26] proposed an accessibility model of emotion recall that suggests that the "emotional experience can neither be stored nor retrieved" [26, p. 935], but rather, during recall, it is inferred from contextual details that are retrieved from episodic memory. Thus, an increase in one's ability to recall contextual details from episodic memory will also lead to an increase in the validity of retrospective self-reporting on experience.

The question that is raised is: how can we best support the recall of episodic memories through external memory cues? Tulving [32] suggests that episodic memories are memories of *who*, *what*, *where* and *when*. As such, recall can be

improved if one is presented with cues about the people involved in an event (*e.g.,* social context cues), the content of the event (*e.g.,* visual cues such as Sensecam photos), the location and the time of the event [1].

Below we attempt to review existing empirical evidence on for each of these types of cues. One has to note, however, that much of empirical evidence in lifelogging relate to memories that do not lie in the near past. Thus, differences may be found across such lifelogging studies and our study.

Visual cues

Episodic memories are dominated by visual imagery and, as such, visual cues have been found to be particularly effective in triggering episodic memories [10, 32]. Not only are they rich in information, visual cues are also *configural*, in that "the objects represented in a visual image are represented in relation to each other and because of this visual images may maximize the amount of information they contain" [10].

Lee and Dey [21] performed a qualitative inquiry into the content of Sensecam photos and identified four different types of cues that trigger recollection: *persons* (e.g., daughter, grandchildren), *objects* (e.g., birthday cake, stained glass window), *places* (e.g., the façade of a visited store, the dining room), and *actions* (e.g., driving home, playing the piano).

Visual cues have driven much of the recent research in lifelogging, notably through Sensecam, a wearable camera that takes 2-3 photos per minute through the person's field perspective. The effectiveness of Sensecam photos in cueing episodic memories even several months following the capture has called the attention of researchers in autobiographical memories as it suggests that "the many episodic memories formed in a typical day are not themselves lost and can often be accessed using sequences of photographs from the camera" [10]. A plausible strength of Sensecam is that, being worn around a person's neck, it maintains the person's original perspective. Research in autobiographical memories has found that visual episodic memories that maintain the person's original perspective are more strongly associated with recollective experience than observer memories, ones that keep a third-person perspective and the person may see themselves in the memory [23].

Given the effectiveness of Sensecam photos in the context of ligelogging (e.g., Kalnikaite et al. [16] Sellen et al., [28]), *Footprint Tracker* may benefit from the presence of such field-perspective photos. Different from prior research on Sensecam, *Footprint Tracker* will cover shorter periods of time in the range of hours to few days. Not only have these episodic memories been subjected to less decay, recent visual episodic memories are more likely to maintain a field-perspective than ones that lie further in the past [25] and thus visual stimuli that maintain the person's original perspective are expected to have a greater capacity to lead to recollective experience [10].

Location cues

Contrary to visual cues that have a rich capacity to trigger detailed information from episodic memory, location cues have been found to support remembering through enabling inferences from established patterns of behavior rather than a true recollection of the event [16]; yet, location cues may in some cases augment visual stimuli leading to the recall of more events and more contextual details surrounding the event [16]. However, locations need to vary significantly if they are to provide information distinct from visual cues [36]. Whitaker et al. [36] reported that, as many Sensecam images in their study were taken at work or at home, this reduced the distinctiveness of location information.

Recent research has also revealed a number of ways for augmenting the information presented by location cues. For instance, Venkatanathan [34] proposed the idea of *trajectory reminders* (i.e., augmenting each location encounter with locations visited before and after). Their study, in the area of location sharing preferences, revealed that trajectory reminders increased participants' test-retest reliability in estimating their preferences for disclosing a given location encounter or not.

Temporal cues

Time is central to episodic memory as "the episodic memory system keeps track of the temporal order of occurrence of personal events" [32] and episodic information is clustered together as episodes by "binding information within an episode to a common temporal context" [12, p. 223].

Temporal cues have been widely employed in retrospective interviewing where recalling the particular time or day in which an event took place assists in setting the context and recalling temporally surrounding events.

Temporal information may have two distinct representations, an *exact* one, where people have been found to use a number of schemes such as day-of-the-week and time-of-the-day that are hierarchically organized, and a *symbolic* one, that stores the temporal order of related events.

Crucial to the design of lifelogging applications such as *Footprint Tracker* is supporting the user in maintaining an awareness of the *temporal context* of each event as temporally aligned events have greater probability of cueing episodic details from each other. One could leverage upon the *primacy* and *recency* effects, the phenomena that people can better recall episodes that lie first or last in a series [11]. For instance, Barsalou [1] asked people to recall their experiences during the last summer. Most participants started in the beginning and proceeded in a chronological order. Conversely, Whitten and Leonard [37] showed that retrieving names of teachers from early schooling years was

more accurate when starting with later years and moving to earlier ones, rather than vice versa [37].

These principles have also been used in retrospective diary methods. For instance, the Day Reconstruction Method [15] asks participants to list their daily activities as a continuous list of episodes which is assumed to form stronger temporal links across the distinct experiences. *iScale* [18] employs graphing as a way to impose a forward temporal order in the reconstruction of one's experiences. This was found to increase the number of reported experiences, the average amount of details in each reported experience, as well as the test-retest reliability of recalled information.

Social context cues

The people we encounter in face-to-face and mediated interactions have for long been assumed to be one of the most effective cues for triggering episodic and autobiographic memories. Lee and Dey [21], in their analysis of what elements within Sensecam photos provide best triggers for memory, found people to be often associated with vivid recollections.

Lifelogging systems have for long attempted to capture the social context of one's daily activities, starting with even one of the very early examples of prosthetic memory, Lamming's and Flynn's 'Forget-me-not' [20] that captured all phone call activities as well as face-to-face contacts through physical pair-wise proximity inference.

However, while many systems have pursued this goal, there is limited empirical knowledge on whether social context does indeed play a significant role in triggering episodic memories, as well as how the different representations of it affect it's impact over remembering. For instance, one might question whether mediated social interactions such as phone calls and email exchanges have equal power with face-to-face contacts and whether they function in a similar way; or, what elements of a social interaction would be best triggers for a remembrance, being it the other person's facial expressions, the content of the discussion, or others. All these are crucial to the design of lifelogging tools that support remembering and reflection over recent events as well as ones lying further in our past.

FOOTPRINT TRACKER

A concern that has been raised both by lifelogging researchers [28] as well as researchers in memory psychology [10] relates to the burden that is imposed on users of lifelogging applications as they are typically presented with an overabundance of photos, location encounters and others. While prior work has focused on eliminating redundant information and clustering visual or location lifelogs into likely coherent events, with *Footprint Tracker* we took an alternative path: that of providing navigation mechanisms. Given that data in *Footprint Tracker* span a limited period of time in one's recent past, from the past few hours to past few days, users are expected to have a better remembrance of the events than what is typical in most lifelogging applications.

Motivated by recent work [18], *Footprint Tracker* places substantial emphasis on the temporal representation of each event and attempts to impose a chronological order in reconstruction. It currently supports three kinds of lifelogs: a) *visual* data as captured from the Vicon Revue camera (a.k.a. Sensecam), b) *location* data representing both significant location (e.g., ones that an individual has spent more than 5 minutes in a 50 meter radius), as well as transitions, c) *social context* data reflected in SMS and calls made and received throughout the day.

The interface is split down into two main sections:

a) *timeline pane* (bottom) highlighting the presence of visual data (green indicates presence of visual data, and blue an absence of data), location data (solid green indicates no transitions, dashed green represents transition and blue indicates an absence of location data), and context data (solid green represents a phone call and its respective duration, a red line targets incoming messages, while blue means lack of cell phone data). These four bars are interactive, allowing users to choose (and adjust) a period of time where they wish to view data (see figure 1),

b) *data pane* (top) that depicts location, visual, temporal data. By selecting a period of time in the *timeline pane*, images and the respective GPS location will be loaded. The *location pane* presents both significant locations (i.e., ones that an individual has spent more than 5 minutes in a 50 meter radius) as well as *trajectory reminders* (i.e., locations visited before and after the current location) [34].

A video of *Footprint Tracker* has been uploaded to the ACM Portal and may assist the reader in understanding how users interact with it.

FIELD STUDY OF FOOTPRINT TRACKER

Method

Participants

A total of 12 participants took part in the study (8 male, median age = 25 , s.d = 3.5). Participants where rewarded with a 10€ voucher for their time. None of our participants' suffered previous memory impairment.

Lifelogging apparatus

Throughout the study, participants carried two devices: a) a Vicon Revue camera (a.k.a. Sensecam) that was worn around their neck and continuously captured photos every 20-30 seconds, and, b) a Google Android device that was continuously logging their location and social context cues (*i.e.,* all exchanged SMS and phone calls). Nine of the 12 participants owned an android smartphone. The remaining three were provided with a Google Nexus device. As we were concerned about how this would impact their usage, especially with respect to SMS, we made sure all

personal data (SMS and contacts) were properly transferred, and users were familiar to the interface and the keyboard.

Study design and procedure

To inquire into the effect of lifelogs as triggers for episodic memories, we followed a simple factorial design with two conditions, *with cues* and *without cues*. All participants joined both conditions; the order was counterbalanced. In the without cues condition participants used a stripped-down version of *Footprint Tracker* where visual, location and social context cues were removed. This provided a close resemblance to the Day Reconstruction Method [15]. The study consisted of two phases: a *capture* and a *recall* phase.

Capture phase. In the capture phase participants carried the lifelogging apparatus for approximately six hours, between noon and 6pm. The length of capturing was limited by the battery lifetime of the Vicon Revue camera. Each condition took place over one day.

Recall phase. At the beginning of each next day, participants uploaded all data on the server and came to our lab for the recall phase. Participants were asked to "review their day as a continuous series of episodes and, when they recall a particular event, to add this along with a description". Participants were asked to think aloud [33] throughout their full recall process. While we considered retrospective think-aloud protocols [13], our pilot study revealed that participants had trouble recalling their thinking process, while think aloud during interaction with *Footprint Tracker* was something that participants did naturally, even when not being asked to do so. A researcher was always next to the participant for any problems encountered with the interface but refrained from engaging in eye contact or discussion with the participant.

Throughout the recall phase, we use a Tobii TX300 Eye Tracker and computed the total *gaze duration* (i.e., the sum of the duration for all fixations within an area of interest) and the *number of visits* (i.e., the number of fixations in an area of interest that have followed a fixation outside the area) over the different areas of interest representing the four types of cues.

Results

Number, duration and richness of recalled events

Participants recalled a total of 274 events throughout the study. When using *Footprint tracker* with cues, participants elicited significantly more events than when using *Footprint Tracker* without cues (paired samples t-test: $t(11)= 8.8,\ p<0.01$; With-cues: mean = 15.7, SD=1.9; Without-cues: mean = 7.2; SD = 1.6).

While participants recalled more events in the presence of cues, the average duration of recalled events in the with-cues condition was significantly lower than those elicited in the without-cues condition ($t(256) = 8.7,\ p<0.01$; With-cues: mean = 16.3 minutes, SD = 17.8; Without-cues: mean = 42.6 minutes, SD = 30.5).

Together, these two analyses suggest that when provided with external memory cues, users are better able to reconstruct their days in greater detail, thus resulting to more events of shorter duration. As one participant noted:

> *"[P10] Viewing my day with pictures is much more accurate than without these.... Without pictures I would remember I was working and shopping, but not [what I was doing] in-between."*

To inquire in the richness of participants' recalls, we submitted these to a content analysis using a predefined schema derived from Lee and Dey [21]. This schema identifies four different details in recalled events: *persons*, *objects*, *places* and *actions*. Each recalled event was characterized by the number of references (zero, one or more) to each of the four details.

Reports elicited in the with-cues condition were found to have significantly more references to objects, $t(249)=0.8$, $p<0.01$, but not with respect to people, $t(249)=1.4$, $p=0.45$, places, $t(249)=3.2$, $p=0.17$, and actions, $t(249)=0.092$, $p=0.93$.

Why and how lifelogs mediate memory

To inquire into the relative dominance of the four types of lifelogs on participants' recall process, we analyzed the total gaze duration and number of visits for each of the four types of logs over the 10 seconds preceding the insertion of a new event by the participant. We estimated this 10-second span empirically through an examination of users' verbal protocol, seeking to understand when participants shift to the cognitive processing of a new event.

An analysis of variance with gaze duration as dependent variable and type of cue (visual, location, temporal and social context) as independent variable revealed a significant main effect of cue on duration, $F(3,752) = 722.74,\ p<0.01,\ h^2_p=0.8$. Post-hoc tests using the Bonferroni correction revealed significant differences at the

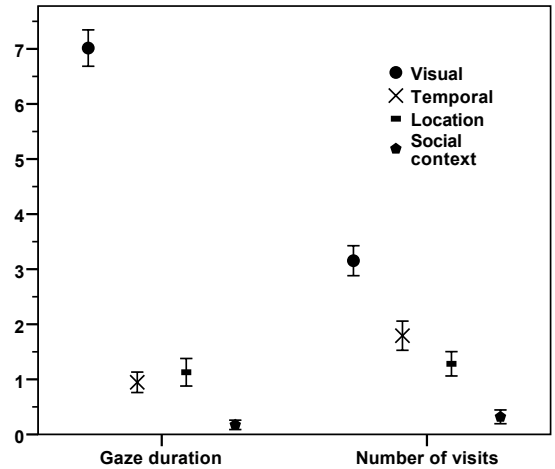

Figure 2. Average gaze duration within the 10-second span and number of visits for the four types of lifelogs.

p<0.01 level for all cue pairs except location and temporal cues.

Visual cues were found to attract substantially more gaze attention than other cues with an average duration of 7.0 seconds (SD=2.3) and 3.2 visits (SD=1.9) on average throughout the 10-sec period. This was due to a number of reasons. First, visual lifelogs provided rich cues of one's daily activities such as the locations and people encountered, but also contextual information surrounding these encounters, for example:

> *"[P2] The images helped me allot, because they allowed me to see the surroundings of my day. From people to buildings, I was able to see everything. I could see my entire day through the images".*

Figure 3. Heatmap from a 10-second capture preceding the insertion of one event.

Visual lifelogs not only provided rich cues of one's daily activities and experiences, but also cues, such as people or objects, that had a particular significance to the activity under recall (e.g., a person one was talking to, a word document one was working on). Similar to Lee and Dey [21], we found that people one encountered and objects he or she interacted with would often trigger memories:

> *"[P3] Oh, the m-ITI guys came with us to have something to eat, we talked about Evangelos app and the best phone to program", "[P3] Oh, I took this paper upstairs to show you my database – I wouldn't have remembered what I was doing If I hadn't seen these pictures!", "[P6] Most of the events that I add are related to interactions with other people. People kept getting memories out of me", "[P6] Oh yeah, I had decided to go home, I can see. I waited 5 minutes for Ijaz but he decided not to come, so I left home",*

Often, however, visual logs would lack personally significant cues, leaving users confused about their whereabouts or their actions at a certain time of the day:

> *"[P1] I am walking, but I'm not sure where", "[P2] I don't recall much from this. You know, it's just a continuous reading/writing cycle. It doesn't have distinct elements, it doesn't have interactions with people, that's why it makes it difficult for me."*

Interestingly, we observed that, rather frequently, especially in the case of repetitive behaviors, users' recollections would suffer from the telescoping effect [30], i.e., the incorrect recall of the time of occurrence of an event:

> "*[*P9*]* what was I doing at the university again?", "*[*P7*]* There is something wrong in the pictures. I didn't go to the toilet at this time, I went later... oh, ok, I was walking down the corridor towards my office (…) oh, this must have been the first time I went to the bathroom, oh yes it's right!"

We found that users would attempt to recover from these errors in two ways. First, they would feed-forward or backward in time seeking for alternative visual cues, e.g.: *"[P2] I don't recall what we were talking about... ...[new photos appear] so, we were discussing the feedback that Carlos gave us of our inDesign project."* Second, they would focus on other sources of information to verify their assumptions or acquire new information. For instance, often, participants would use location logs to verify the exact location when this could not be inferred from visual logs, or would look at the social context logs to read the exact SMS they send at a given point in time, as their mobile phone was displayed in the Sensecam photos.

Temporal cues appeared to be the second most frequently gazed part of the interface. While their average gaze duration of 1.0 second (SD = 1.3) was not significantly different from that on location cues, *t(187)=1.2, n.s.*, participants exhibited a significantly higher number of visits to temporal cues (mean=1.8, SD=1.8) than to location cues (mean=1.3, SD=1.5, t(187)=3.1, p<0.01). It thus appears that participants spent more time each time they gazed at their location, yet, they gazed more frequently at the temporal cues.

Temporal cues supported recollection as they provided a temporal context to each recall, e.g., *"[P10] Time helped me situate myself in my day, and think of what I was doing at a certain time"*. Especially in cases of stable daily patterns, users could reliably infer their location or activity simply by looking at time information, e.g., *"[P8] I have a scheduled timetable – I come to work everyday at 09:00 and leave at 18:00 – Time helped me place myself in a certain activity, since I knew, for example that I came to work around 08:45"*.

Temporal cues supported not only inferences about the ongoing activity but also about upcoming activities as people could tap either into their recent memories, e.g., *"[P2] I'm not sure what Carlos was showing me, but I know that I was in a hurry to go home, since it was my lunch time"*, or into their own habits and external knowledge such as the open hours of a supermarket, *"[P10] I must have been close to leaving home because I go shopping every day. It's half past seven and Pingo Doce closes at eight"*.

Last, rather often, we found that temporal cues supported reasoning when misjudgments were made based on other cues**.** For instance, as the following participant was gazing at Sensecam photos, he commented: *"[P12] I think this is the time when I went to the Social Security..."*. The participant then gazed at time and commented: *"...oh no, it's still 11 o'clock, I only went there around 17o'clock"*.

Location cues. Location cues were the third most frequently gazed part of the interface with an average gaze duration of 1.1 seconds (SD = 1.8) and an average of 1.3 visits per recall (SD=1.5).

We found that, in most cases, participants were able to infer their location from the Sensecam photos. Participants would most often gaze at the location information either when Sensecam photos did not provide this information, for instance when driving or walking in an unfamiliar setting, or when they wanted to confirm their location inference from Sensecam photos, for example:

> *"[P3] Location information wasn't very important to me because I was able to determine where I was through the photos. I could know more precisely where I was, but I didn't need it to recall what I did. I even forgot about the map, and then I would every once in a while look at it for curiosity", "[P8] …. was only useful when I was driving. I knew where I was going, but I checked the Location to know where I was at a precise moment. It didn't help me remember anything, just used it for curiosity."*

Interestingly, we found that trajectory reminders (i.e., locations visited before and after the current location) supported location inference when participants could not infer their current location from Sensecam photos (i.e., locations visited before and after the current location). As an example, participant 6 viewed a set of images of himself leaving a building, gazed at the location cues and commented: *"I can see, I was going to Marias, to have a break"*.

Last, one has to note the different role of the two representations of location information. The map would display the current location along with locations visited before and after (*i.e.,* trajectory reminders). The location bar would instead, highlight whether location information was present and whether one was static, or in transition. We noticed that some users used the location bar as a way to navigate through different events of their day, by jumping to timeframes where they were in transition, e.g., *"[P6] So, I started moving over here, lets see where I was going"*.

Social context cues displayed little relevance to participants' recall process. Interestingly, participants commented that this was due to the lack of novelty of the information as they could have access to this information throughout the day, contrary to visual and location cues, e.g., *"[P1] I didn't pay much attention to it since it's something I have access on a daily basis on my phone"*.

However, we found instances where participants used social context cues in unexpected ways. For example, when seeing himself dial a number on his mobile phone, one participant opted to confirm the person he was calling at the moment: *"[P3] … here I was calling Cátia"*. In other cases, these social contacts acted as milestones in one's past day, cueing the recall of temporally proximal events:

> *"[P1] My aunt Sofia called me. I picked her up and we went to Martin's day care center", "[P4] there was a situation in which I received a message asking if I wanted to have lunch, which helped me remember that I was having lunch around that time*".

Users' preferences

An analysis of users' ratings of the two different versions of *Footprint Tracker* (with and without cues) revealed significant differences both in *perceived usefulness* (with-cues: mean=4.3, SD=1.4; without-cues: mean=2.6, SD=1.1; $t(11)=6.1$, $p<0.01$) and in *perceived enjoyability* (with-cues: mean=5.1, SD=0.8; without-cues: mean=3.3, SD=1.3; $t(11)=6.1$, $p<0.01$).

To inquire into what aspects of *Footprint Tracker* were most effective in assisting the recall process, we analyzed users' ranks of the four types of cues (i.e., visual, location, temporal and social context) for the two questions: their ability to recall: (a) the *emotions* associated with reported events, and (b) *episodic details* relating to reported events (see figure 2).

Pair-wise Wilcoxon signed rank tests revealed that visual cues were ranked significantly higher than all other types of cues in their ability to assist the recall of *episodic details* (*e.g.,* visual-temporal: $Z=-3.1$, $p<0.01$) and *emotions* (*e.g.,* visual-temporal: $Z=-3.1$, $p<0.01$). All remaining differences were non-significant.

Visual cues were ranked first by all 12 participants in recalling episodic details and by 11 out of the 12 participants in recalling emotions. During the final

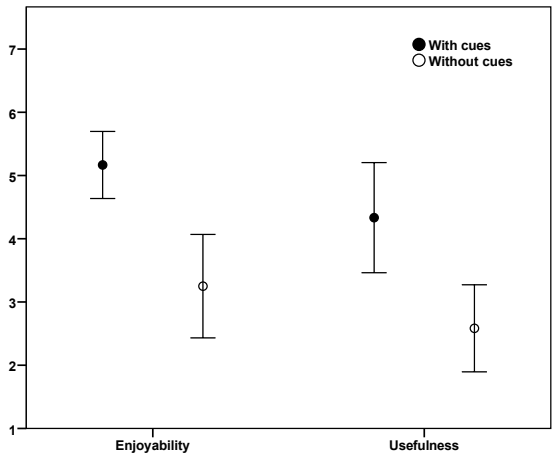

Figure 4. Participants' ratings on the perceived usefulness and enjoyability of the two versions of Footprint Tracker

interviews, participants' highlighted the relevance of visual cues in recalling episodic details surrounding daily events and associated emotions, often cued by different objects or people displayed in the picture:

"[P1] Yes, I think pictures made me remember a lot of things, and a lot of details I didn't remember. I consider it to be important because I could see not only where I was, but also who was with me.", "[P6] ...I remember that I felt bad after eating the pumpkin… the other information can't help me remember how I was feeling".

Temporal cues were the second highest ranked cue, with seven participants ranking these second most effective in recalling episodic details and five in recalling emotions. The effectiveness of temporal cues as a trigger to episodic details was largely tied to participants' stable daily routines as well as the recency of the events under recall, as participants in our study recalled daily activities that took place the day before, unlike most studies in the area of lifelogging. Contrary to the recall of episodic details, temporal cues were not as effective in the recall of emotions: *"[P7] Time didn't help me remember what I was feeling. Time by itself didn't trigger any emotions".*

Location cues were the third highest ranked cue with seven participants ranking these third most effective in recalling episodic details and five in recalling emotions. Participants commented that location cues were often redundant as they could infer these through the Sensecam photos, and that location cues were only relevant in transitions, e.g., *"[P9] Time was more important than Location because I didn't move around a lot. I was mostly inside the Tecnopolo building".*

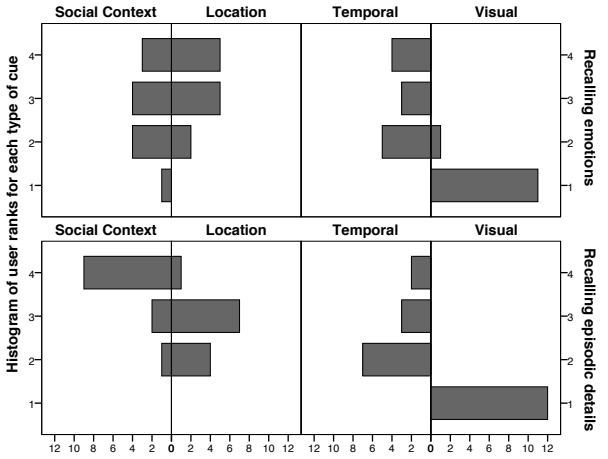

Figure 5. Histogram of user ranks for the four different types of cues, in assisting the recall of emotions or episodic details. Visual cues are ranked first, followed by location and temporal cues. Social context cues rank last in assisting the recall of episodic details, but perform better in the recall of emotions.

Social context cues were the lowest ranked type of cue when it comes to assisting the recall of episodic details, with 9 out of the 12 users ranking these last. Interestingly, participants commented that this was partly due the lack of *novelty* of the information, as they had access to these SMS and phone calls any time they wanted, e.g., *"[P1] I didn't pay much attention to it since it's something I have access on a daily basis on my phone".* Yet, interestingly, users judged social context as a more effective cue when it comes to recalling the *emotions* in contrast to recalling episodic details (Mann-Whitney U Test: Z=0.08, n.s.), with one participant ranking social-context as the most effective cue in recalling emotions: *"[P3] The time and location didn't help me remember how I was feeling. But this (social contextual data) helped me remember how I was feeling because I remembered I was talking to Cátia and what we said to each other."*

DISCUSSION AND CONCLUSION

Overall, our findings suggest that supporting participants with lifelogging during diary studies increases their ability to recall and reflect upon their daily activities and experiences. One has to note that our control condition (without cues) was a close resemblance of the Day Reconstruction Method [15] as *Footprint tracker* provides an awareness of the temporal context of recalled events, following the same principles of DRM. Thus, the difference of lifelog-supported recall to organic memory is expected to be even greater.

As Tulving [32] suggested, episodic memories are memories of *who*, *what*, *where* and *when*, and recall can be improved if one is presented with cues about the people involved in an event, the content of the event, the location and the time of the event [1].

The question we raised was: which of the four type cues may best trigger episodic memories of recent events? Our findings corroborate but also extend recent research in the area of lifelogging. Different from lifelogging research, our study addressed recalls over recent events (*i.e.,* past day).

In line with our expectations and prior literature [8, 10, 16, 28], we found that visual cues, attracted the most visual attention (in terms of gaze duration as well as number of visits), were often the starting point of participants' recollections, and were rated by all participants as the best memory trigger. Contrary, location cues were rarely the focal point of a recall – out of the 189 recalled events, only for 14 of them (7%) location cues attracted more visual attention than visual cues, and most participants ranked them as the third most effective memory trigger. These findings corroborate Kalnikaite's et al. [16] finding that visual cues support true recollections as opposed to location cues that support remembering through inference. Yet, in Kalnikaite's et al. [16] study visual cues had a surprising limited power as memory triggers in terms of the number of recalled events. We believe this is mostly due to the gradual loss of accessibility of visual episodic memories [10]. Thus, it appears that while location cues may support recall over

prolonged periods of time, visual cues are more valuable as triggers for memories of recent events.

While social context cues appeared to be the least significant cues in aiding recall, one has to be aware of a number of limitations of *Footprint Tracker* in representing social context. First, we found social context often to be inferred from Sensecam photos. Such, face-to-face encounters were more frequent and perhaps more meaningful and memorable than mediated social interactions as captured by *Footprint Tracker*. Secondly, one could argue that social media such as Facebook and Twitter may better reflect individuals' social interactions than calls and short messages exchanged from their mobile phones. While we agree that *Footprint Tracker* should be extended to cover those media, we wish to distinguish between directed communication, such as a phone call or a private communication between two contacts on Facebook, and non-directed communication such as public announcements on Facebook and Twitter. Prior work has shown directed communication exchanges to have stronger impact on individuals' psychological state than non-directed ones [7]. Thus one would expect directed communications to be more personally relevant and consequently act as stronger memory cues of our daily social interactions. In our limited sample, participants had an average of 4.8 (median=4) incoming calls and 3.4 (median=3) incoming SMS per day, which is comparable to the average number of directed exchanges on FB [7]. This suggests that social media such as Facebook complement traditional media in our social interactions, but do not necessarily dominate, as one might initially expect.

Yet, while visual cues were the dominant memory trigger, we found that other cues assisted recall in unexpected ways. First, participants used other cues, such as location and social context cues, to confirm inferences made through the Sensecam photos, thus supporting a feeling of knowing that has been shown to enhance recall [14]. Secondly, participants used other cues to resolve conflicting information, such as a conflict between the location inferred through Sensecam photos and one displays in the location pane, as well as to correct misjudgments such as the temporal misplacement of recalled events [30]. Third, participants other cues for maintaining an awareness of the temporal order of events under recall, for instance, through temporal cues, or through social context cues acting as milestones in one's day.

While lifelogging brings the potential of decreasing the obtrusiveness of self-reporting in diary studies, one could argue that the continuous capturing of cues for memory may entail a number of drawbacks over traditional diary methods where participants self-select when and what to capture.

First, traditional diary studies may benefit from the so-called *self-selection effect* [29], according to which when people self-select the events to report, they forget them at a slower rate than events that have been recorded randomly, likely due to increased attention in self-selected events. However, empirical findings in the self-selection effect are mixed. For instance, Sellen et al. [28] found that passively captured photos through Sensecam provided better cues for recall than actively captured ones. Future research should attempt to study whether Sellen's et al. [28] findings hold in recently captured photos as in the context of diary studies.

Second, one has to be aware that wearing a Sensecam is not completely unobtrusive. Prior work has revealed that Sensecam induces a feeling of being under surveillance to participants and raises privacy concerns on others [24]. In our study we found Sensecam to affect behaviors in two ways. Some participants reported avoiding public locations when wearing it while others used it to initiate social interactions. Further work is needed to understand the impact Sensecam has on the findings of diary studies.

Third, continuous capturing produces an overabundance of data and may impose a substantial burden to the participant during review and recollection, a challenge that is known to the lifelogging community [3, 21, 31]. Research in lifelogging has attempted to overcome this challenge through two primary approaches: *data elimination,* whereby some of the data are permanently deleted, or *data clustering*, whereby data are grouped by their similarity. Both approaches face unique challenges. For instance, users are not willing to delete personal data [22, 35], and this deletion may also lead to the risk of missing potentially significant cues. Contrary, clustering needs accurate processing while being flexible enough for user intervention, thus imposing substantial algorithmic challenges.

We argue that, in the context of diary studies where recollections happen over limited periods of time, such as a few days to weeks, complicated algorithmic approaches are not required. Instead, we argue for approaches that facilitate navigation through data. With *Footprint Tracker* we pursued two such navigational cues: *micro-views* and *temporal context*.

Micro-views are small snippets of information that provide an overview of the presence and nature of lifelogs at a single glance. In *Footprint Tracker*, this was supported through the timeline pane, indicating the presence or not of visual cues (i.e., Sensecam photos), the presence and nature (i.e., static or in transition) of location cues, as well as social context cues.

Temporal context was supported in *Footprint Tracker* through two representations of time: an *exact* (top right part of the interface) which provided a quick glance to temporal information, and a *symbolic* one, which supported an awareness of the temporal order of events under recall. While the relevance of temporal information has been noted by researchers of episodic and autobiographical memory

[10], lifelogging tools have yet to unleash the full potential of time as a trigger for memories.

Concluding, our study aimed at providing a first inquiry in the potential of lifelogging as an approach to increasing the *unobtrusiveness* and *realism* of diary studies. While our study provided promising results for the plausibility of lifelogs as memory cues in this context, we have yet to address several challenges if we are to establish lifelogging as viable methodological paradigm within diary studies. Most importantly, future work should focus on a direct comparison of self-selected and automatically captured media and the practical benefits of lifelogging in actual design and evaluation tasks.

REFERENCES

1. Barsalou, L. W. (1988). The content and organization of autobiographical memories. In: Neisser, U., Winograd, E.(Eds.), Remembering reconsidered: Ecological and traditional approaches to the study of memory, 193-243.
2. Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.
3. Blighe, M., & O'Connor, N. E. (2008). MyPlaces: detecting important settings in a visual diary. In ACM CIVR, 195-204.
4. Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. Annual review of psychology, 54 (1), 579-616.
5. Brandt, J., Weiss, N., & Klemmer, S. R. (2007). txt 4 l8r: lowering the burden for diary studies under mobile conditions. In CHI'07 extended abstracts, 2303-2308.
6. Brown, B. A., Sellen, A. J., & O'hara, K. P. (2000). A diary study of information capture in working life. In Proc. of CHI 2000, 438-445.
7. Burke, M., Marlow, C., & Lento, T. (2010). Social network activity and social well-being. In Proc. of CHI'08, 1909-1912.
8. Carter, S., & Mankoff, J. (2005). When participants do the capturing: the role of media in diary studies. CHI'05, 899-908.
9. Consolvo, S., & Walker, M. (2003). Using the experience sampling method to evaluate ubicomp applications. Pervasive Computing, IEEE, 2(2), 24-31.
10. Conway, M. A. (2009). Episodic memories. Neuropsychologia, 47(11), 2305-2313.
11. Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. Journal of experimental psychology, 54(3), 180.
12. Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. Psychological review, 119(2), 223.
13. Guan, Z., Lee, S., Cuddihy, E., and Ramey, J. (2006) The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In Proc. of CHI '06, 1253-1262.
14. Hart, J. T. (1965). Memory and the feeling-of-knowing experience. Journal of educational psychology, 56(4), 208.
15. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. Science, 306 (5702), 1776-1780.
16. Kalnikaite, V., Sellen, A., Whittaker, S., & Kirk, D. (2010). Now let me see where i was: understanding how lifelogs mediate memory. In Proc. of CHI 2010, 2045-2054.
17. Karapanos, E. (2012). Experience Sampling, Day Reconstruction, what's next? Towards technology-assisted reconstruction, M-ITI internal report.
18. Karapanos, E., Martens, J.B., & Hassenzahl, M. (2012) Reconstructing Experiences with iScale. International Journal of Human-Computer Studies, 70 (11), 849-865.
19. Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J.B. (2009). User experience over time: an initial framework. In Proc. of CHI '09, 729-738.
20. Lamming, M. and Flynn, M. (1994) 'Forget-me-not' Intimate Computing in Support of Human Memory, In Proc. Of *FRIEND21, 1994.*
21. Lee, M. L., & Dey, A. K. (2007). Providing good memory cues for people with episodic memory impairment. In Proc. of ACM ASSETS 2009, 131-138.
22. Marshall C. (2007) How People Manage Personal Information over a Lifetime, In W. Jones and J. Teevan, (Eds.), Personal Information Management, 57-75.
23. Mcisaac, H.K., Eich, Eric. (2002), Vantage point in episodic memory, *Psychonomic Bulletin & Review*, 9 (1), 146-150.
24. Nguyen, D. H., Marcu, G., Hayes, G. R., Truong, K. N., Scott, J., Langheinrich, M., & Roduner, C. (2009). Encountering Sensecam: personal recording technologies in everyday life. In Ubicomp'09, 165-174.
25. Robinson, J. A., & Swanson, K. L. (1993). Field and observer modes of remembering. Memory, 1(3), 169-184.
26. Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. Psychological bulletin, 128(6), 935.
27. Scollon, C.N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. Journal of Happiness Studies, 4, 5–34.
28. Sellen, A. J., & Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. CACM, 53(5), 70-77.
29. Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of experimental Psychology: Human learning and Memory, 4(6), 592.
30. Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988). Telescoping in dating naturally occurring events. Memory & Cognition, 16(5), 461-468.
31. Truong, K. N., & Hayes, G. R. (2009). Ubiquitous computing for capture and access. Foundations and Trends in Human-Computer Interaction, 2(2), 95-171.
32. Tulving, E. (1984). Precis of Elements of Episodic Memory. Behavioural and Brain Sciences 7 (2), 223–68.
33. Van Den Haak, M., De Jong, M., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & Information Technology, 22(5), 339-351.
34. Venkatanathan, J., Ferreira, D., Benisch, M., Lin, J., Karapanos, E., Kostakos, V., Sadeh, N., Toch, E. (2011). Improving users' consistency when recalling location sharing preferences. In Proc. of INTERACT 2011, 380-387.
35. Whittaker, S., Bergman O., Clough P. (2010) Easy on that trigger dad: a study of long term family photo retrieval. Personal and Ubiquitous Computing, 14 (1), 31-43.
36. Whittaker, S., Kalnikaité, V., Petrelli, D., Sellen, A., Villar, N., Bergman, O., Clough, P., & Brockmeier, J. (2012). Socio-technical Lifelogging: Deriving Design Principles for a Future Proof Digital Past. Human–Computer Interaction, 27, 37–62.
37. Whitten, W.B., Leonard, J.M. (1981), Directed search through autobiographical memory, Memory & Cognition, 566-579.