ORIGINAL ARTICLE

EmoSnaps: a mobile application for emotion recall from facial expressions

Evangelos Niforatos · Evangelos Karapanos

Abstract We introduce EmoSnaps, a mobile application that captures unobtrusively pictures of one's facial expressions throughout the day and uses them for later recall of her momentary emotions. We describe two field studies that employ EmoSnaps in an attempt to investigate if and how individuals and their relevant others infer emotions from self-face and familiar face pictures, respectively. Study 1 contrasted users' recalled emotions as inferred from EmoSnaps' self-face pictures to ground truth data as derived from Experience Sampling. 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. Study 2 assessed EmoSnaps' ability to capture users' experiences while interacting with different mobile apps. The study revealed systematic variations in users' emotions while interacting with different categories of mobile apps (such as productivity and entertainment), social networking services, as well as direct social communications through phone calls and instant messaging, but also diurnal and weekly patterns of happiness as inferred from EmoSnaps' self-face pictures. All in all, the results of both studies provided us with confidence over the validity of self-face pictures captured through EmoSnaps as memory cues for emotion recall, and the effectiveness of the EmoSnaps tool in measuring users' momentary experiences.

E. Niforatos (✉)
Faculty of Informatics, University of Lugano, Lugano, Switzerland
e-mail: evangelos.niforatos@usi.ch

E. Karapanos
Madeira Interactive Technologies Institute, Funchal, Portugal
e-mail: ekarapanos@uma.pt

1 Introduction

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, the Experience Sampling Method (ESM) [1]. ESM is often considered as the "gold standard" of in situ measurement [8] as it samples experiences and behaviors right at the moment of their occurrence, thus reducing memory and social biases in self-reporting. However, ESM also entails significant drawbacks, such as disrupting a users' current activity and imposing an additional reporting burden [16]. Motivated by these drawbacks, Kahneman and colleagues proposed the Day Reconstruction Method (DRM) [8], 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. DRM has been found to provide a reasonably good approximation to Experience Sampling data [8], and the method has been well adopted also in the HCI community.

In our line of research, we attempt to contribute toward a next step in the field of momentary assessment that of technology-assisted reconstruction (TAR) [10]. TAR consists of passively logging users' behaviors throughout the day with mobile sensor technology and employing these data to assist the reconstruction 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 her momentary emotions. In the remainder of the article, we first present the theoretical basis on emotion and memory, followed by an elaboration on the EmoSnaps solution. Next, we present two studies: (a) a two-week-long deployment of EmoSnaps that inquired into *if* and *how* self-face pictures assist the reconstruction of momentary emotions, and (b) an extensive deployment that attempted to assess the value of EmoSnaps in evaluating a wider set of mobile interactions and their associated emotions.

1.1 Emotions, facial expressions, and memory

Emotions are so tightly connected to facial expressions that one could even question whether there can be emotion without facial expression [2]. 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 [3]. In particular, when it comes to happiness, research has revealed that humans can accurately recognize the emotion in 96.4 and 89.2 % of the times in Western and non-Western cultures, respectively [4].

Algorithmic techniques in emotion recognition have flourished [5–7] 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 [8] on "Self-Cam," 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 [9] 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.

1.2 Memories of 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 [10]. 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. Tulving later introduced 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 [11].

Based on Tulving's distinction of episodic and semantic memory, Robinson and Clore [12] proposed an accessibility model of emotion, according to which "an emotional experience can neither be stored nor retrieved," but it can be inferred from contextual details springing from 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 on the event and then 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.

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* [11]. 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 [13]. Social interactions have been proven to be one of the most effective cues for triggering episodic and autobiographic memories. For example, Lee and Dey [14] used SenseCam [15] 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 [16]. Another interesting approach in supporting episodic memory recall is presented in [17]. 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.

Episodic memory has a primarily visual nature, and as such, visual cues have been proven to be exceptionally effective in assisting the recall process [11, 16]. The reason why visual cues are so effective in triggering memories lies in the so-called configural nature of visual images and the ability of represented objects to relate to each other, maximizing the information they contain [18]. Despite the fact that location cues lack the immediateness that visual cues induce during recall, they have been found to implicitly support remembering through enabling

inferences from established patterns of behavior rather than a true recollection of an event [19]. However, location cues need to vary significantly to single-handedly support episodic recall [20]. Time also plays a major role in recall since it is the main driver according to which personal events are registered in episodic memory [11]. 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 temporally adjacent events [16]. A more detailed description available in [16] summarizes the effectiveness of the aforementioned cues to trigger episodic and auto-biographic memories. The current work examines the potential of self-face pictures to facilitate affective recall from episodic memory and, therefore, aims to improve the validity of retrospective self-reporting on experience.

1.3 EmoSnaps

EmoSnaps is a mobile application that captures pictures of one's face using the front-facing camera of a mobile device (Fig. 1). EmoSnaps employs event-driven sampling, where predefined events, such as "screen unlock," "phone call answer," "SMS sent," and application launches, trigger a picture capturing. In our first study, we limited the event-driven sampling 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 using the front-facing camera. EmoSnaps is able to capture a picture

Fig. 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

within 300–500 ms, adding only minimal interaction delay in the process and hence being almost transparent to the user. Following a successful picture capturing, no sampling occurs for the next 5 min.

2 Study 1: 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:

(a) Are participants able to recognize their emotions on self-face pictures captured during mobile device usage? Given prior literature [3, 4], 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 recognize their emotions in self-face pictures, the question is how do they do so? We can think of at least two ways (Fig. 2). The first assumes that individuals will recognize their emotions through their facial expressions in the self-face picture [4, 7]. In contrast, the second option assumes that individuals will use cues of the picture to recall episodic memories (e.g., where they were, what they were doing, who they were with) and, based on this information, are 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 cues from episodic memory [21]. If self-face pictures are recent and contain information that may cue episodic memories, participants could as well
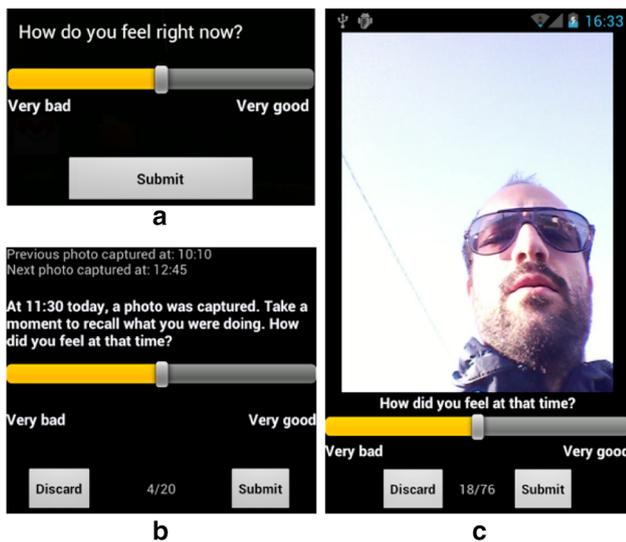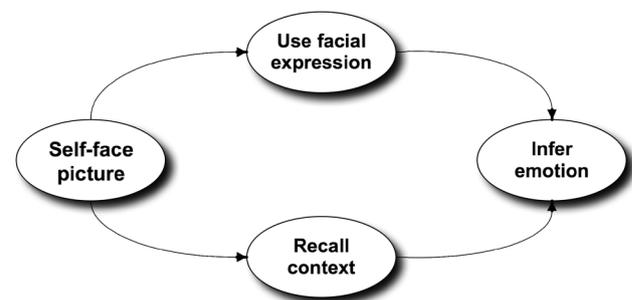
Fig. 2 If people accurately recognize their emotions, how do they do so? Inferring emotions directly from facial expressions, or recalling episodic memories and drawing upon this knowledge to infer emotion?

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 [22]. Others have shown, however, that familiarity may make a difference by improving the accuracy in recognizing emotions [23]. Given these results, we expect familiar others to be better able to infer the individual's emotions from her facial expressions.

2.1 Study design

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

2.1.1 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 (Fig. 1c). Pictures are presented in chronological sequence as this has been proven to enhance the reconstruction of episodic cues [24]. Thus, we assume participants in this condition to have access to both approaches of emotion inference: *recognition* and *reconstruction*.

2.1.2 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 (Fig. 1b). No actual pictures are shown but are 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 [25]. 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.

2.1.3 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 interface as in the Photo-Day condition (Fig. 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.

2.1.4 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 participants' facial expressions. Relevant others used the same interface as in Photo-Day and Photo-Week reconstructions (Fig. 1c) with the pictures being displayed in random order.

2.2 Measures

Motivated by the previous work in the field of self-report on psychological well-being [26–30], we designed a simple interface (Fig. 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. 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 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" (Fig. 3).
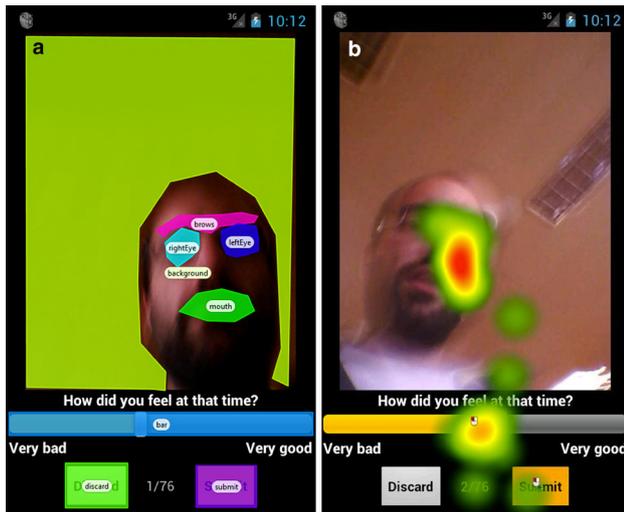
Fig. 3 **a** Clustering Areas of Interest (AOIs) for eye tracking analysis for each picture. **b** Heat map produced by summarizing gaze behavior on a self-face picture

Fig. 4 Interface used for performing Retrospective Think Aloud (RTA) sessions

During eye tracking analysis, two metrics were measured: visit count and total visit duration. 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 interface when asked to infer their emotion, four metrics were derived: "Photo duration" (in seconds) describes the overall time taken to evaluate a picture, "events number" holds the total number of bar touches per picture, and "total events duration" (in milliseconds) describes the total duration of all bar-touch events observed per picture. Finally, "total delta" describes the total distance covered by the bar cursor during reconstruction. Retrospective Think Aloud (RTA) 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 cues (Photo-Day, Photo-Week, and Photo-Relevant). For this purpose, an RTA protocol was formed mainly questioning the rationale behind emotion inference.

2.3 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.

2.4 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. Each of the seven participants was given a Nexus S mobile device with the EmoSnaps application preinstalled and was asked to use it as her 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 [26] (Fig. 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 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 (Fig. 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) [31] session using as cues the two captured videos combined and presented in one interface (Fig. 4).

3 Results

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

3.1 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. For the remaining 29.4 %, they clicked on 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 her 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).
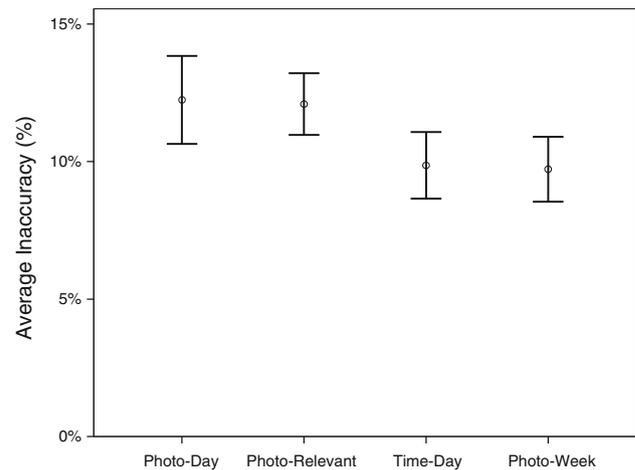
Fig. 5 Z-transformed average inaccuracy per condition. Photo-Week was significantly more accurate compared to Photo-Day and Photo-Relevant

3.2 Context recall versus facial expression use

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 ($F(3,998) = 4.553$, $p < .01$, $h_p^2 = 0.014$). 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, $p < .05$) (Fig. 5). However, no other significant effects were found.

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 ($F(2,72) = 4.251$, $p < .05$, $h_p^2 = 0.106$). 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, $p < .05$) (Fig. 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

---

[1] Z-transformation was applied to normalize the distance Δ between ESM and reconstruction ratings.
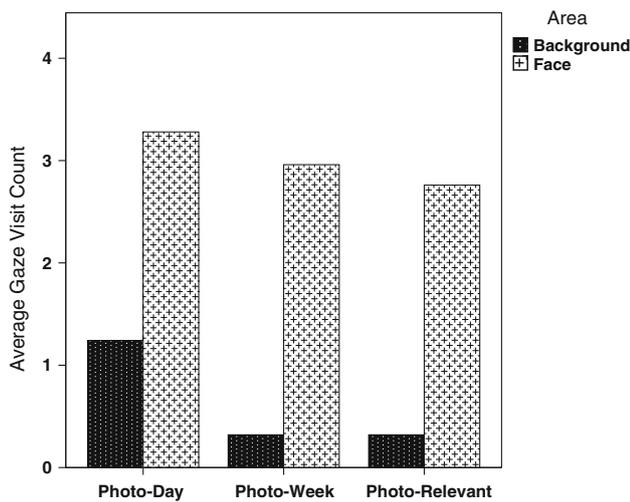
Fig. 6 Average Gaze Visit Count (average number of times eye gaze visits an area) per condition for Background and Face areas in a face picture. Face was proven the most dominant in eye gaze visits

picture to infer their emotion than they did one week after a picture was captured.

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:

> "[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 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.

3.3 Relevant versus 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, $M = 9.722$, $SD = 9.629$, $p < .05$) (Fig. 5). 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, $M = 11.333$, $SD = 6.813$, $p < .05$) (Fig. 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, $M = 0.320$, $SD = 0.556$, $p < .05$) (Fig. 6). This effect is complemented by
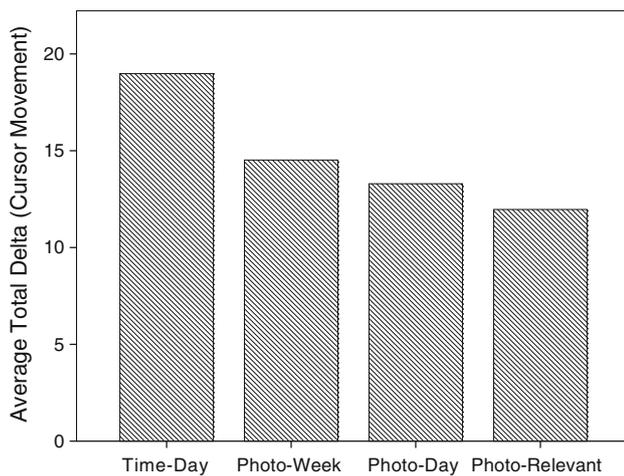
Fig. 7 Average distance covered by cursor per condition (0–99) during evaluation. Time-Day entailed the highest bar cursor movement

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 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. 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 (Fig. 5). This contradicts our initial hypothesis, 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 capturing 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 that the relevant others of an individual are better able to infer the individual's emotions from her 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 [5]. 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 (Fig. 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) [25]. 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 [11] and therefore emotion. Unfortunately, no qualitative data are available for this condition since it was initially designed as a control condition. Additional findings concern errant ways according to 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 to capture naturalistic behaviors, while promoting meaningful sampling, we decided to employ transparent face picture capturing 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 that one performs when she wants to access her mobile device. 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 capturing 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."

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 recognition process. An alternative possible explanation could be a learning effect in the Photo-Week condition, as participants were more familiar with the reconstruction interface since they used it before in the Photo-Day condition.

5 Study 2: Deploying EmoSnaps in the wild

While the first study provided promising results for the effectiveness of EmoSnaps, 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 apps. 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.

5.1 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: (a) screen unlock, (b) call answer, call end, or outgoing call, (c) SMS sent/read, (d) application launches, and (e) 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 needed 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 sampling 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?

5.2 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 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?" [26–28] 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.

5.4 Research questions

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 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 [32–34]. For example, morning hours are related to lower levels of happiness and higher levels of annoyance and anger [32]. 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 [33]. 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 [35], 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 [36]. 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 [37]. 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 [37].

6 Results

Before presenting the results of the study, we define as events the mobile interactions that fall into the following five broad categories: (a) screen unlock, (b) call answer, call end (incoming/outgoing), and placing a call, (c) SMS sent/read, (d) application launches, and (e) 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 [38, 39], 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 (0.9 %): Astro, Mega, Dropbox, etc.
- Image viewing (0.8 %): Gallery, Album, Infinite view, etc.
- Entertainment (0.5 %): 9gag, Angry Birds, Simpsons, and other mobile games
- Google Play (0.5 %): Android vending
- Video Playing (0.4 %): YouTube, Android Video Player, MX Tech video player
- Security (0.3 %): Avast, Clean Master, etc.
- Weather (0.3 %): AccuWeather, Genie Widget, Weather Widget
- News (0.3 %): Pulse, Flipboard, etc.
- Text Reading and Editing (0.3 %): Adobe Reader, Polaris viewer, Think Droid
- Music (0.2 %): Shazam, Jango mobile, etc.
- New App Installed (0.1 %): Android Package Installer

6.1 Discard rates

All in all, participants were able to infer emotions for approximately 50 % ($N = 1{,}477$) of their self-face pictures; 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. First, during the second study, a greater range of events was captured, some of which do not imply an appropriate posture. Second, the increased sampling 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.

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,
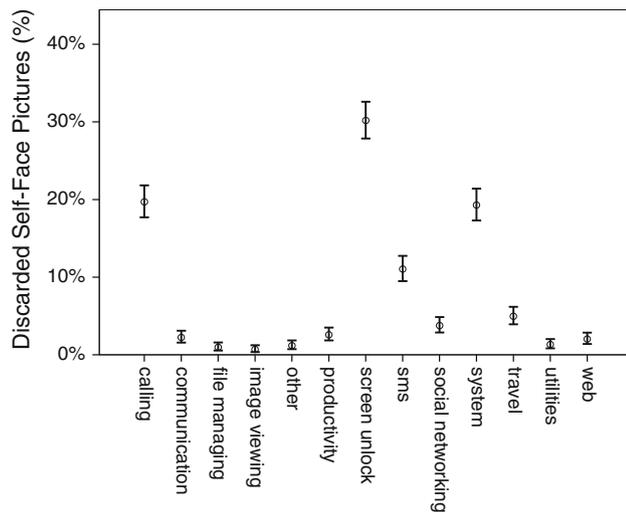
Fig. 8 Number of discarded self-face pictures generated for each event category. Categories with equal or lower than 0.5 % of total occurrence are excluded

whereas in the previous study, the same event displayed a significantly lower discard rate (29.4 %) ($\chi^2(1, 1,462) = 71.654$, $p < .001$) (Fig. 8).

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 of interaction these categories imply. Incorrect posture of the face in front of the mobile device or insufficient capturing time affects the overall quality of self-face pictures being captured (Fig. 9). 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 (Fig. 9).

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 nighttime. 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 (Fig. 10a). 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 nighttime, which has a strong influence on the quality of the captured self-face pictures (Fig. 9).

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

Fig. 9 Some examples of pictures that participants discarded. In the first one from the left, the luminosity is high; in the second one, the camera focus time is insufficient; and in the third one, the participant's face is not fully included in the picture
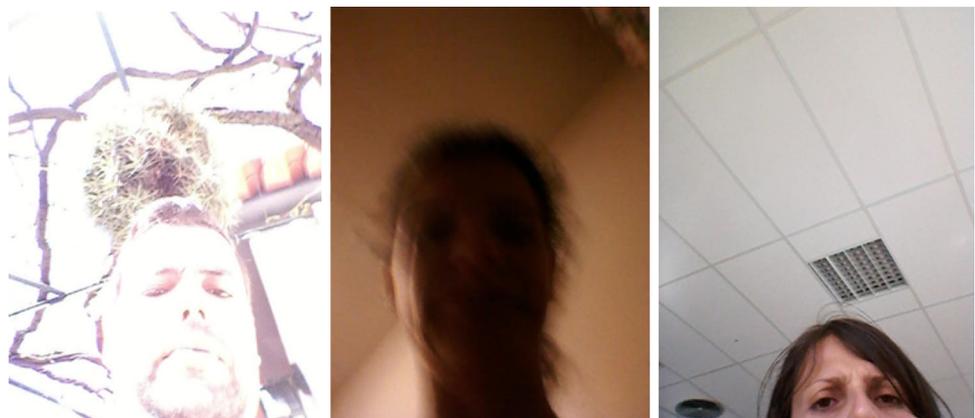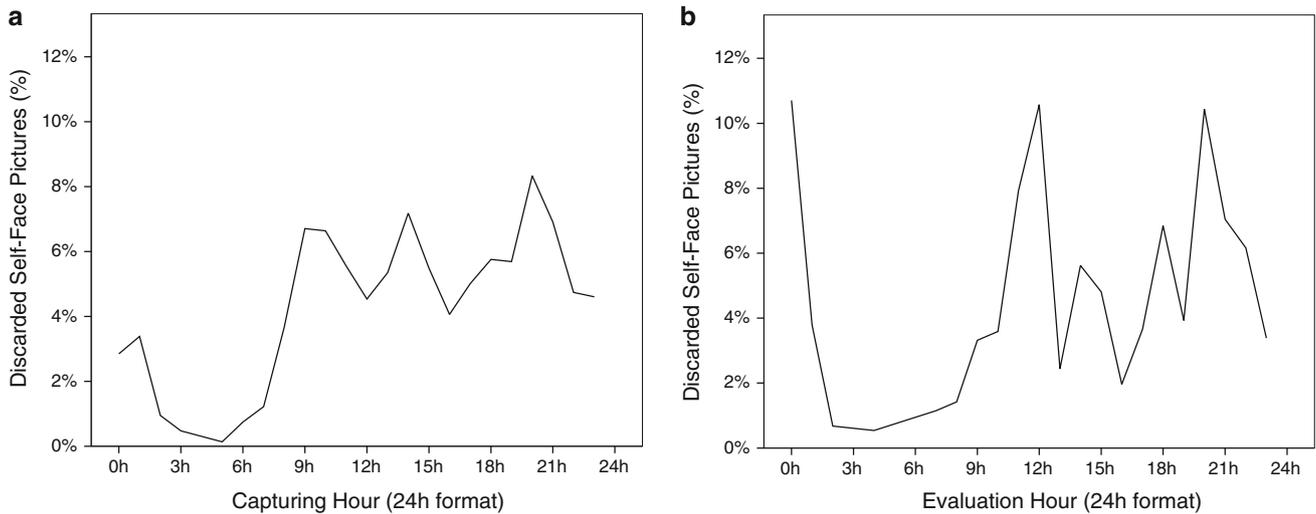
Fig. 10 Ratio of discarded self-face pictures in relation to **a** the hour they were captured and **b** the hour they were evaluated

of the day had the highest discard rate, up to 80 % at 23:00 (Fig. 10b). A Pearson's chi-square analysis between discarded and rated distributions for the above timeframes 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 [32], 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 %). 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 on the weekend [40].

6.2 Happiness fluctuation 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. While no pattern can be observed during weekends, during weekdays, happiness seems to display a low in the early hours of day with a constant increase over the course of the day (Fig. 11a).

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 ($F(201,285) = 3.741$, $p < .001$, $h_p^2 = 0.056$) (Fig. 11a), but not for the weekend. Post hoc tests using the Bonferroni correction revealed that self-face pictures particularly captured at 08:00 ($M = 2.688$, $SD = 0.143$, $p < .05$) 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 ($M = 4.25$, $SD = 0.286$) were found to be significantly happier than pictures captured at 08:00, 10:00 ($M = 3.136$, $SD = 0.072$, $p < .05$), 12:00 ($M = 3.037$, $SD = 0.089$, $p < .05$), and 14:00 ($M = 3.143$, $SD = 0.082$, $p < .05$). No other significant effects were found. These results are in line with existing findings in the psychology of well-being, 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 [32].

Systematic effects are also observed in the variation of happiness over the course of the week (Fig. 12). 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 ($F(61,476) = 7.68$, $p < .05$, $h_p^2 = 0.03$) (Fig. 12). Post hoc tests using the Bonferroni correction revealed that pictures evaluated on Tuesday ($M = 3.465$, $SD = 0.887$, $p < .05$), Wednesday ($M = 3.326$, $SD = 0.839$, $p < .05$), Thursday ($M = 3.353$, $SD = 0.675$, $p < .05$), Friday (M $= 3.192$, $SD = 0.892$, $p < .05$), and Sunday ($M = 3.473$, $SD = 0.844$, $p < .05$) were rated significantly happier than pictures evaluated on Monday ($M = 2.83$, $SD = 0.591$). In addition, pictures evaluated on Sunday were also rated significantly happier
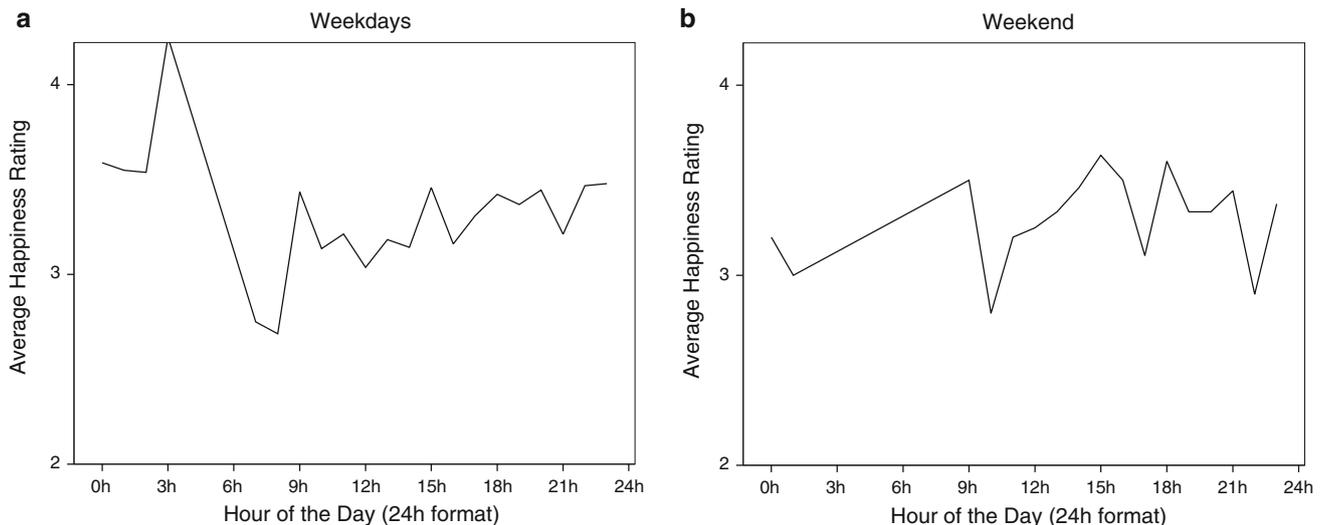
Fig. 11 Average happiness rating fluctuation observed during **a** weekdays and **b** weekends, as reported based on self-face pictures captured
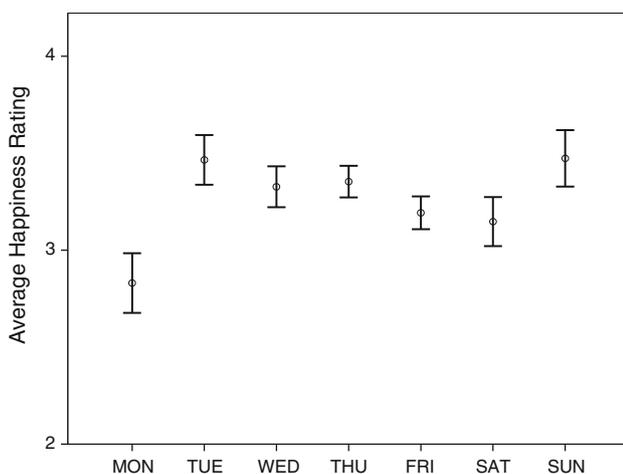
Fig. 12 Average happiness rating per evaluation day for all self-face pictures

than pictures evaluated on Saturday ($M = 3.147$, SD $= 0.818$, $p < .05$) and Friday ($M = 3.192$, SD $= 0.892$, $p < .05$), but no additional significant effects were found. These results corroborate the well-known "Blue Monday" and "Weekend" psychological effects [32, 33]. Overall, EmoSnaps seems able to capture the daily and weekly variation of happiness, as reflected in the current psychological literature.

6.3 Happiness fluctuation 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 0.5 % 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 ($F(12,1476) = 2.278$, $p < .001$, $h_p^2 = 0.027$) (Fig. 13). Given prior research, one would expect that events relating 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 [40, 41]. Surprisingly though, we found the exact opposite with respect to the category "Calling" (Fig. 13). Post hoc tests using the Bonferroni correction revealed that self-face pictures captured through events falling into "Productivity" ($M = 3.689$, SD $= 0.76$ $p < .01$), screen unlock ($M = 3.348$, SD $= 0.84$, $p < .005$), "Social Networking" ($M = 3.43$, SD $= 0.745$, $p < .05$), and "System" ($M = 3.321$, SD $= 0.725$, $p < .05$) categories were rated significantly happier than self-face pictures captured when "Calling" ($M = 3.046$, SD $= 0.914$) events occurred. 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 [41]. 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 [34–36]. 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
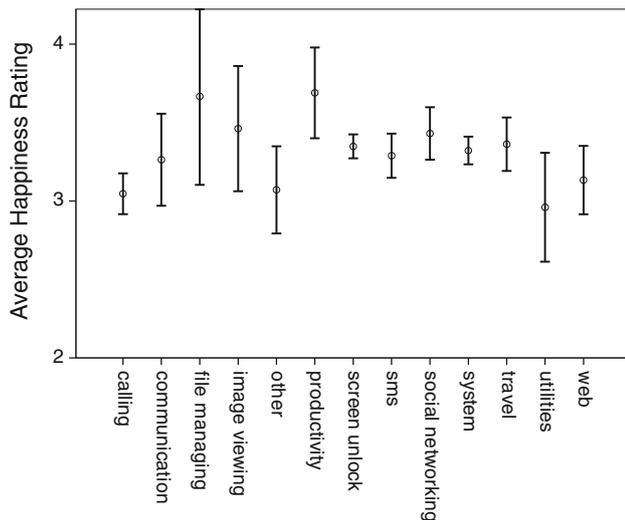
Fig. 13 Average happiness per event category as reported based on self-face pictures. Categories with equal or lower than 0.5 % of total occurrence are excluded

found to produce significantly happier self-face pictures than "Calling" did. 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 [42], and we thus 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 that social networking applications will produce happier self-face pictures.

Interestingly, the "screen unlock" event displayed very small standard deviation (SD = 0.84, N = 465) compared to other event categories (Fig. 13), 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 ($M = 3.283$). 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 ($t(464) = 2.422$, $M = 0.094$, SD = 0.084, $p < .05$). 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.

6.4 Habits emergence

Recent research has shown that a standard mobile use session lasts approximately less than a minute [37, 38]. In pursuit of revealing additional behavioral insights, we clustered all occurring events in (overlapping) timeframes 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. 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 variables revealed a significant main effect for the number of preceding events ($F(8, 1430) = 3.001$, $p < .005$, $h_p^2 = 0.017$). 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 = 0.847) than pictures of events that followed after 3 consecutive events ($M = 3.02$, SD = 0.855, $p < .005$) within a 2-min period (Fig. 14), 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. However, the same analysis for happiness rating and subsequent events revealed no significant main effect ($F(8, 1430) = 0.510$, $p > .05$, $h_p^2 = 0.003$). At this point, the top five most occurring 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
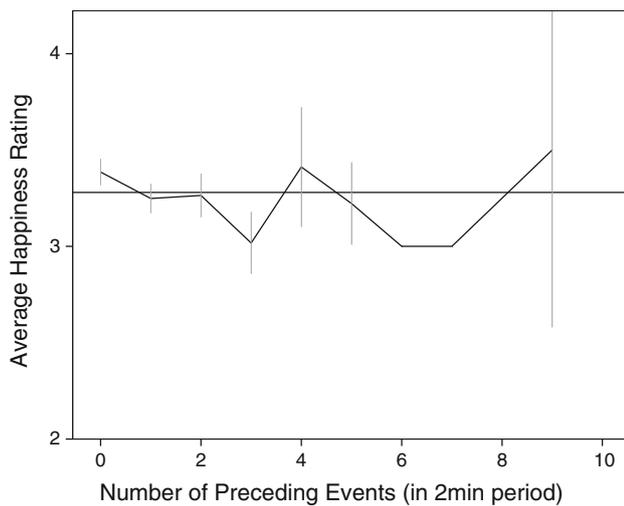
Fig. 14 Average happiness rating per number of preceding events within 2-min period. Pictures of events with no preceding event were rated significantly happier than pictures of events with 3 preceding events

Fig. 15 Screen unlock events rate in relation to number of events they entailed, within a period of 2 min

the type of the event category on the number of preceding ($F(12, 2857) = 92.986$, $p < .001$, $h_p^2 = 0.281$) and the number of subsequent ($F(122,857) = 23.594$, $p < .001$, $h_p^2 = 0.09$) events. Not surprisingly, post hoc tests using the Bonferroni correction revealed that the screen unlock event systematically displayed the minimum number of preceding events ($M = 0.06$, SD $= 0.239$, $p < .001$) 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 event categories, apart from "File Managing" category ($M = 0.81$, SD $= 0.849$, $p > .05$). However, for the 22.1 % of the times that participants unlocked their mobile device, no event was detected for the next 2 min (Fig. 15). 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 (Fig. 15). 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 $= 0.914$), strong indications arise for a checking habit formation on behalf of the participants [37]. 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 [37, 38]. 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.).

7 Discussion

Overall, this work corroborated findings from studies on psychological well-being, which demonstrates that Emo-Snaps 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. 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 $= 0.835$). 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 [32, 33].

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 [41]. 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 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 might be the increased feeling of control that these types of mobile interactions induce on individuals [42].

Interestingly, our findings confirm prior insights into the habitual use of mobile devices [37, 38]. 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 [37], 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, 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 ($M = 3.283$), when less than 2 h had elapsed between capturing and evaluation (Fig. 16). 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 passed between capturing. A global minimum was observed when 15 h had passed between capturing and evaluation. The displayed variation on reported happiness
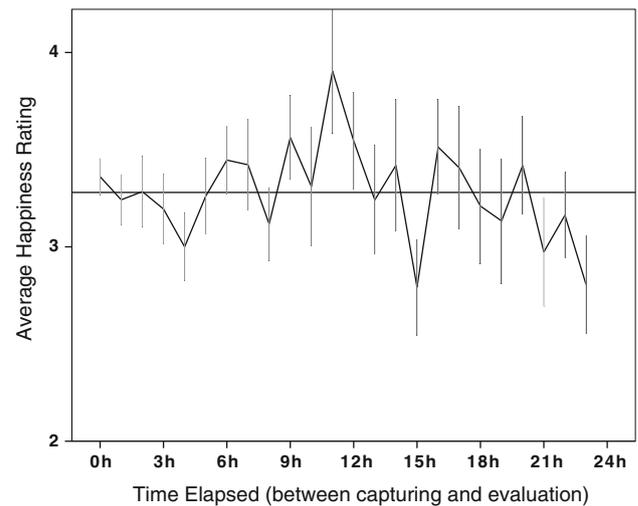
Fig. 16 Average happiness rating in relation to time elapsed between capturing and evaluation

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 [11].

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," "Traveling," 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 [32]. 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.

8 Overall discussion

Both studies aimed at assessing the effectiveness of EmoSnaps as a lightweight tool for measuring users' experiences with mobile applications. The first 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 feelings 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 recall, yet individuals are likely to attempt to recall details from episodic memory and infer their emotions from these, thus introducing memory biases (such as 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.

The first study was very informative as it demonstrated that using EmoSnaps, individuals (a) were able to infer their moment-to-moment emotions remarkably reliably, and (b) even more interestingly, the most effective path to performing this task was through 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 validity of self-face pictures captured through EmoSnaps, as memory cues for emotion recall, and the effectiveness of the EmoSnaps tool in measuring users' momentary experiences.

9 Conclusions and future work

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 behaviors and experiences. In this paper, we proposed 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, sampling is almost invisible to the user, while reconstruction is enhanced by the self-face pictures, so individuals do not rely merely on their memory.

We reported two 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 Emo-Snaps 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 [37]. 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.

Overall, the results suggest that EmoSnaps can be a viable approach to Technology-Assisted Reconstruction [43]. Users were able to recall emotions based upon their facial expressions, 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. Future work will focus on exploring new mobile interactions for capturing self-face pictures of a mobile user and how this could enhance the experience reconstruction for application prototype evaluation.

References

1. Larson R, Csikszentmihalyi M (1983) The experience sampling method. New Direct Methodol Soc Behav Sci
2. Ekman P (1993) Facial expression and emotion. Am Psychol 48:384
3. Ortony A, Turner TJ (1990) What's basic about basic emotions? Psychol Rev 97:315
4. Russell JA (1994) Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. Psychol Bull 115:102
5. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22:1424–1445
6. Cohen I, Sebe N, Garg A, Chen LS, Huang TS (2003) Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Underst 91:160–187
7. Azcarate A, Hageloh F, van de Sande K, Valenti R (2005) Automatic facial emotion recognition. Universiteit van Amsterdam
8. Teeters A, El Kaliouby R, Picard R (2006) Self-Cam: feedback from what would be your social partner. p 138
9. Gruebler A, Suzuki K (2010) Measurement of distal EMG signals using a wearable device for reading facial expressions. pp 4594–4597
10. Bartlett FC (1995) Remembering: a study in experimental and social psychology. Cambridge University Press, Cambridge
11. Tulving E (1984) Precis of elements of episodic memory. Behav Brain Sci 7:223–268
12. Robinson MD, Clore GL (2002) Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 128:934–960. doi:10.1037//0033-2909.128.6.934
13. Barsalou LW (1988) The content and organization of autobiographical memories. Cambridge University Press, Cambridge
14. Lee ML, Dey AK (2008) Lifelogging memory appliance for people with episodic memory impairment. In: Proceedings of the 10th international conference on Ubiquitous computing. ACM, New York, NY, USA, pp 44–53
15. Hodges S, Williams L, Berry E, Izadi S, Srinivasan J, Butler A, Smyth G, Kapur N, Wood K (2006) SenseCam: a retrospective memory aid. UbiComp 2006: Ubiquitous Computing 177–193
16. Gouveia R, Karapanos E (2013) Footprint tracker: supporting diary studies with Lifelogging. In: Proceedings of CHI 2013
17. Sas C, Fratczak T, Rees M, Gellersen H, Kalnikaite V, Coman A, 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
18. Conway MA (2009) Episodic memories. Neuropsychologia 47:2305–2313
19. Kalnikaite V, Sellen A, Whittaker S, Kirk D (2010) Now let me see where i was: understanding how lifelogs mediate memory. In: Proceedings of the 28th international conference on human factors in computing systems, pp 2045–2054
20. Whittaker S, Kalnikaitė 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 Comput Interact 27:37–62
21. Karapanos E, Martens JB, Hassenzahl M (2009) Reconstructing experiences through sketching. arXiv preprint arXiv:0912.5343
22. Young AW, McWeeny KH, Hay DC, Ellis AW (1986) Matching familiar and unfamiliar faces on identity and expression. Psychol Res 48:63–68
23. Kenny DA, Acitelli LK (2001) Accuracy and bias in the perception of the partner in a close relationship. J Pers Soc Psychol 80:439
24. Anderson SJ, Conway MA (1993) Investigating the structure of autobiographical memories. J Exp Psychol Learn Mem Cogn 19:1178
25. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA (2004) A survey method for characterizing daily life experience: the day reconstruction method. Science 306:1776–1780. doi:10.1126/science.1103572
26. Conner TS, Reid KA (2012) Effects of intensive mobile happiness reporting in daily life. Soc Psychol Pers Sci 3:315–323
27. MacKerron G, 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
28. Killingsworth MA, Gilbert DT (2010) A wandering mind is an unhappy mind. Science 330:932. doi:10.1126/science.1192439
29. Isaacs E, Konrad A, Walendowski A, Lennig T, Hollis V, Whittaker S (2013) Echoes from the past: how technology mediated reflection improves well-being. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1071–1080
30. Sanches P, Höök K, Vaara E, Weymann C, Bylund M, Ferreira P, Peira N, Sjölinder M (2010) Mind the body!: designing a mobile stress management application encouraging personal reflection. In: Proceedings of the 8th ACM conference on designing interactive systems. ACM, pp 47–56
31. Guan Z, Lee S, Cuddihy E, Ramey J (2006) The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1253–1262
32. Stone AA, Smyth JM, Pickering T, Schwartz J (1996) Daily mood variability: form of diurnal patterns and determinants of diurnal patterns. J Appl Soc Psychol 26:1286–1305
33. Ryan RM, Bernstein JH, Brown KW (2010) Weekends, work, and well-being: psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. J Soc Clin Psychol 29:95–122
34. Rystrom DS, Benson ED (1989) Investor psychology and the day-of-the-week effect, pp 75–78
35. Jin B, Peña JF (2010) Mobile communication in romantic relationships: mobile phone use, relational uncertainty, love, commitment, and attachment styles. Commun Rep 23:39–51. doi:10.1080/08934211003598742
36. Green MC, Hilken J, Friedman H, Grossman K, Gasiewskj J, Adler R, Sabini J (2005) Communication via instant messenger: short- and long-term effects. J Appl Soc Psychol 35:445–462

37. Oulasvirta A, Rattenbury T, Ma L, Raita E (2012) Habits make smartphone use more pervasive. Pers Ubiquit Comput 16:105–114

38. Böhmer M, Hecht B, Schöning J, Krüger A, Bauer G (2011) Falling asleep with angry birds, Facebook and kindle: a large scale study on mobile application usage. In: Proceedings of the 13th international conference on human computer interaction with mobile devices and services. pp 47–56

39. Ickin S, Wac K, Fiedler M, Janowski L, Hong J-H, Dey AK (2012) Factors influencing quality of experience of commonly used mobile applications. Commun Mag 50:48–56

40. Verkasalo H (2008) Contextual patterns in mobile service usage. Pers Ubiquit Comput 13:331–342. doi:10.1007/s00779-008-0197-0

41. De Guzman ES, Sharmin M, Bailey BP (2007) Should I call now? Understanding what context is considered when deciding whether to initiate remote communication via mobile devices. Proc Graph Interface 2007:143–150

42. Larson R (1989) Is feeling "in control" related to happiness in daily life? Psychol Rep 64:775–784

43. Karapanos E (2011) Experience sampling, day reconstruction, what's next? Towards technology-assisted reconstruction. M-ITI, Funchal