Do Life-Logging Technologies Support Memory for the Past? An Experimental Study Using SenseCam

Abigail Sellen, Andrew Fogg, Mike Aitken*, Steve Hodges, Carsten Rother and Ken Wood
Microsoft Research Cambridge
7 JJ Thomson Ave, Cambridge, UK, CB3 0FB
{asellen, v-afogg, shodges, carrot, krw}@microsoft.com
*Behavioural & Clinical Neuroscience Institute
Dept. of Psychology, University of Cambridge
mrfa100@hermes.cam.ac.uk

ABSTRACT
We report on the results of a study using SenseCam, a "life-logging" technology in the form of a wearable camera, which aims to capture data about everyday life in order to support people's memory for past, personal events. We find evidence that SenseCam images do facilitate people's ability to connect to their past, but that images do this in different ways. We make a distinction between "remembering" the past, and "knowing" about it, and provide evidence that SenseCam images work differently over time in these capacities. We also compare the efficacy of user-captured images with automatically captured images and discuss the implications of these findings and others for how we conceive of and make claims about life-logging technologies.

Author Keywords
Personal digital archives, life-logging, SenseCam, episodic or autobiographical memory, images, capture

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
The idea of a "life-log" or a personal digital archive is a notion that can be traced back at least 60 years [5]. The vision is that technology will allow us to capture everything that ever happened to us, to record every event we ever experienced and to save every bit of information we have ever touched. Indeed, in recent years, this vision has been given a new lease of life, recent impetus and enthusiasm coming mainly from a number of technological advancements. These include the development of smaller, lighter-weight capture devices and sensors (capturing everything from image, location and ambient sound to heart rate), advances in wireless networking, and massive increases in digital storage capacity making the archiving of huge amounts of personal data possible. Now, as never before, technology offers the possibility of capturing data from everyday life both continuously and unobtrusively.

As a result, a number of new efforts to build systems and devices to support life-logging have emerged in recent years. A case in point is a technology developed in our own laboratory called "SenseCam" [18]. SenseCam is a device containing a camera and embedded sensors worn around a user's neck which automatically takes a series of still images over time as well as capturing other aspects of life events such as ambient light levels, temperature and movement.

In addition to SenseCam, there are many other life-logging systems being developed both in research laboratories and as commercial products. Much of this work is only concerned with recording users' activities in the digital world, focusing on interaction with electronic media such as documents, photos, sounds, videos and so on [e.g., 1, 9, 13, 15]. However, another strand of work is more concerned with recording aspects of life "out there" in the real world, away from the desk. Some of the earliest research in this vein can be traced back to the Active Badge and PARC Tab systems at EuroPARC, Xerox PARC, Olivetti and AT&T [e.g., 20, 27, 32] in which users carried or wore small devices which were tracked within the confines of an area covered by networked sensors. More recent instantiations of this approach can be classified either as *wearables, portables or instrumented environments.* Wearable systems are based mainly on head-mounted still or video cameras [e.g., 17, 24,], or on wearable audio capture devices [30]. Portable systems largely make use of specialized software on existing devices such as PDAs, notebook computers or cell phones [e.g., 21, 22]. Cell phone data have also been used to infer human activity after the fact [10, 19] by analyzing location-based data. An alternative approach is to rely on instrumented environments which capture activity through installed sensors or local wireless networks [e.g., 25].

Many potential benefits have been put forward for such systems, but by far the most common proposition is that, by capturing data about our daily activities, life-logging systems will offer effective support for memory of our own personal past [e.g., 2, 17, 20]. These arguments range from supporting the reliving of or reminiscing about personal events, to more specific functional support for memory

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

including finding lost objects or documents, remembering names, remembering whom you met, details of conversations, and remembering past actions or events. In other words, life-logging applications claim to support a whole range of ways in which we can look back, re-live, re-examine, and search through our past experiences.

Such systems generally make some (usually implicit) assumptions by their very design. One is that the more data captured the better in helping us to look back. So these devices aim to capture and store as much data as possible in the course of daily life. Another is that the more *different kinds* of data captured the better. Thus many systems strive to be both as comprehensive and diverse as possible in the amount and variety of data capture that takes place. And another is that capture of data should occur automatically or at least with minimal effort on the part of the user. Thus data are captured throughout the course of everyday events, and as an incidental by-product of a user's activities.

Naturally, one result of such undertakings is a huge amount of data that must be structured, organized and searched through. Indeed much of the interesting work in this area is grappling with the challenges this presents for users, interface designers and computer scientists [12].

However, in this paper we propose that there are equally important research challenges in substantiating the beneficial claims made about these systems. Specifically, there is little systematic evidence that the data offered up by life-logging systems do in fact support human memory, and little research to help us understand how they might do so. To date, we must rely on anecdotal evidence or conjecture. Indeed the rhetoric in this area often rests on unfounded or even irrational assumptions about what these systems do. An example of this is the claim that life-logging technologies capture day to day "experiences". It is important to point out that these systems, no matter how sophisticated, comprehensive, or diverse in their data capture, do not capture human "experience". Rather such technologies capture a set of cues (data) which we hope will trigger the *remembering* of human experience.

This then begs a number of important questions such as: Do the cues these systems capture really support memory of past experience? If so, how much can people recall about the past given a particular kind of cue? What kinds of cues are best (e.g., images, ambient sound, location)? Does the effectiveness of different kinds of cues deteriorate over time? Does it matter if these cues are automatically captured by the technology or actively captured by the user?

The study which we describe in this paper is a step toward explicating and answering questions such as these. Yet, this is not solely in the service of advancing our theoretical knowledge--it is also of relevance more generally for the development of life-logging technologies. Unless we systematically assess the underlying assumptions of such technologies, we can only make "best guesses" about what they are good for, and how they may be most effectively designed.

RELATED RESEARCH

As we have already alluded to, there is very little in the CHI literature which explores the link between life-logging technologies and human memory. While there are many systems which claim to support memory, most often such claims are not substantiated with anything more than anecdotal evidence.

Exceptions include a study, now 15 years old [11], in which three subjects' memory for events during a day at the office was tested after intervals from one day up to four weeks. Subjects were given both automatically generated text-based diaries and video excerpts of their days and found that while the text diaries did increase recall, video cues were more effective. A more recent study (again using a very small sample size) investigated people's memory for computing events within a one hour time slot [8]. In this case, there was only weak evidence that both video and still images can serve as effective reminders of these past events after one day and one month. Research by Vemuri et al [30] also provides some evidence for the use of audiotaped conversations (in this case, conference talks) in jogging memory, although in this case the memory problems were simulated rather than real. Finally, Carter and Mankoff [6] explored the efficacy of different kinds of media as cues in eliciting recall of everyday events within the context of diary studies. In a study involving 7 participants, they compared the use of photos, ambient sounds and tangible objects in cueing the recall of a day-long festival. Not many details are given with respect to the method and analysis, but in general they found that photos were best for eliciting detail.

Turning to the literature in the psychology of human memory, we face a different problem: it is often difficult to make direct connections from contrived laboratory situations to the kinds of real-world situations we are interested in. Having said that, there is much theoretical work that can be applied.

For example, we can begin with the distinction between "semantic" and "episodic" memory [29]. Semantic memory is the term used to refer to general knowledge of the world (e.g., that a dog has four legs, that the capital of France is Paris). Episodic memory, on the other hand, refers to the capacity to re-experience specific episodes from your past (e.g., feeding your dog last night, or the time you visited the Eiffel Tower). With life logging technology, it is clear that most of the claims that are made have to do with enhancing and exploiting the relationship between captured cues and episodic (or what is sometimes called "autobiographical") memory.

With regard to episodic or autobiographical memory, a further useful distinction is made by Conway [7] who distinguishes the *recollection of specific details of recent past events* (remembering what one perceived or felt at the time of an event) from *more general knowledge about oneself* accumulated over time. The former is about mentally traveling back in time; the latter is more about the "things one knows about oneself" such as the periods in one's life, repeated patterns and habits of activities, and the stories one tells about oneself. Throughout the rest of this paper we will use the term "remember" or "recollect" to refer to the former concept: the mental re-experiencing of the original event, as opposed to the more schematic, general self-knowledge one might bring to bear in a memory task. This distinction will become important later as we shall see.

Aside from theory, the relevant psychological research for our purposes largely comes from a sub-field of memory psychology often called "ecological" or "everyday" memory, a reactionary movement to conventional lab-based memory research encouraging the study of memory in natural contexts. In particular, diary studies in which people (often the researchers themselves) recorded everyday events over long periods of time are of relevance here [4, 23, 31]. We will refer to these specifically in the next section.

EXPERIMENTAL QUESTIONS AND APPROACH

Given the plethora of issues that we have outlined, this study must be viewed as one component of a larger program of work designed to unpack and evaluate the effectiveness of these technologies. Our starting point is to examine the assumption that data such as photographic images can provide effective cues for memory of past experience. With SenseCam, still images form a key part of the data collected. For this reason, we began by focusing on these, and in particular wanted to look more closely at the fact that, with SenseCam, images are taken passively, i.e. without user intervention. This makes SenseCam different from most people's experience of cameras where images are captured intentionally.

In designing this study, our goal was to generate realistic data by having our subjects wear SenseCam in the course of their everyday lives, but we also aimed to test subjects' memory in a controlled laboratory setting. Specifically, for this study, we had four main questions in mind:

(i) Do SenseCam images improve our memory for past personal events above and beyond what we would normally remember?

(ii) In what ways do SenseCam images help us connect with our past? For example, can people even recognize images as their own? Do they really help us recollect (re-experience) past events?

(iii) If SenseCam images are effective cues for memory, how does this change over time?

(iv) Are passively captured SenseCam images better retrieval cues for everyday memory than similar intentionally or actively captured images?

With regard to the first question, psychological studies based on people keeping diaries over weeks [4] and even years [23, 31] have shown the efficacy of various cues such as details of "who", "what", "where" and "when" in evoking recollective memory. While none of these studies used images (only written diary entries), preliminary work with an amnesic patient [3] indicates that SenseCam images may in fact be effective triggers for recall. We therefore did predict that SenseCam images would improve recollection of past events.

With regard to the second question, we were interested in testing memory in at least three ways: to see whether people could distinguish their own images from those of others (in a recognition test); to see whether they could sequentially order their own images; and to see whether they could report events from their past when presented with their own images (a "cued recall" test). At the same time, we designed the study in such a way to distinguish "remembering" or "recollection" of past events from simply "knowing" something had occurred. The ways in which SenseCam images would support these different kinds of memory, and the way this would change over time (our third question) we viewed as exploratory issues.

The fourth question is of interest because if one thinks about the practice of writing diary entries or taking photos, these acts of recording are fundamentally different from SenseCam in that people make choices about which aspects of experience they record. We hypothesized, therefore, that such actively captured records would be more potent memory triggers over time, or, to put it another way, that passively taken images would be weaker triggers than actively taken images to recalling the past. Certainly past research has shown that diary studies in which people self-select events to record forget them at a much slower rate than events recorded randomly [see 4]. Of further relevance here is a well established phenomenon known as the "generation effect" [28] which finds that people are better at remembering materials (usually words) they have generated themselves than materials they are merely presented with. With regard to recognition, we also thought that actively taken pictures would be more meaningful to people and thus easier to recognize and to order than passive images.

METHOD

Subjects

A total of 10 male and 9 female university undergraduates took part in the study, ranging in age from 18 to 22 years. Subjects were paid £50 on completion of the experiment. The choice of undergraduates was deliberate given the claims often made about how life-logging technologies can benefit all of us, not just the elderly or memory-impaired.

Experimental Design

We were interested in three main independent variables: whether or not subjects wore SenseCam ("SenseCam" versus "Control" conditions); whether images taken with SenseCam were automatically taken or user initiated ("Passive" versus "Active" capture modes); and whether memory was tested at a short or a long interval (3 versus 10 days). All of these variables (which we will call "Condition", "Mode", and "Interval") were explored using a within-subjects design, meaning each subject experienced all combinations of these three variables.

Table 1 shows the ordering and schedule of these conditions. Subjects wore the SenseCam for two consecutive days during which it passively captured images. On those same days, subjects were also asked to actively take pictures (details below). Two Control days were considered to be either side of the SenseCam days. One SenseCam day and one Control day were tested at the short interval, the other two days were tested at the long interval.

In order to control for order effects of short and long interval testing, two different orderings were used as a between-subjects factor with subjects randomly assigned to one group or the other (A or B). Note that this design results in some variation in the length of the short and long intervals for the testing of Control days only.

Table 1. The letter 'c' refers to a control day, 's' to a SenseCam day and 't' to a test day. Capital letters denote long-interval capture, control or test days; lowercase letters denote short-interval capture, control or test-days.

	Day								
	1	2	3	4	5	6	…	12	13
Group A	C	S - - -	- - -	- - -	- - -	- - -	- -//- -	→ T	
			s	c - -	- →	t			
Group B	c	s - - -	- - -	- - →	t				
			S	C - -	- - -	- - -	- -//- -	- - -	→ T

Procedure

On the two SenseCam days, subjects were asked to wear a camera for eight hours (10am-6pm) each day collecting the cameras from and returning them to the same location in the mornings and evenings.

On arrival on the first day, subjects were instructed how to use the camera and were told that from time to time it would automatically take pictures. In addition, they were asked to take at least 40 pictures a day (at least 20 before 2 pm, and 20 after 2 pm) by manually pressing a button on the camera. They were instructed to take these photos "as if you are creating a visual journal of your day". Subjects were subsequently asked to return on days 6 and 12 (Group A) or 5 and 13 (Group B) to complete the study.

The procedure for testing at the short (3 day) and long (10 day) intervals was the same. There were two main kinds of tests in a session, recall and recognition, the second of which was fully automated and run on a laptop computer. Each session took approximately an hour to complete:

Recall. These tests were designed to discover which events that occurred on the day in question subjects could recall before and after viewing a given set of images. Each session tested one of the Control days and one of the Sensecam days, dividing up each day into morning or afternoon. Half of the subjects were tested for a Control day first, and half for a SenseCam day first.

Initial free recall: Subjects were first asked to recall as much as they could about the first half of the day in question. They were given two minutes in which to write a description onto a paper worksheet providing detail about "what", "where", "when" and "who" for each remembered event.

In addition to detailing these events, they were asked, for each event, to tick "remember", "know", or "guess" (a common distinction in memory tests (see [14] for a review). It was explained to subjects that rather than a scale of well-remembered to badly-remembered, the three options represented qualitatively different types of memory which we defined as follows:

- *Remember* – We defined this as when an event can be re-experienced in the 'mind's eye', where one can mentally place oneself in the scene described.
- *Know* – This was defined as an event which one infers *must* have occurred that day, perhaps because it was a routine event (e.g. a music lesson on a Tuesday), or perhaps because they 'remember' spending time with someone later in the day, so therefore *must* have spent time with them in the morning as well, even though they are not able to mentally re-experience doing so.
- *Guess* – We suggested the use of this option to allow subjects to fill in events they were uncertain about (perhaps out of a desire to comply with the test protocol when nothing much was remembered).

Viewing and ordering of images: Subjects were then presented with a randomly-ordered series of ten images (see Figure 1). If tested about a SenseCam day, these were their Active or Passive images from one half of that day. Alternatively, if tested about a Control day, they were shown Active or Passive images taken by another person they were experimentally "yoked" to. In this case, this meant another subject also participating in the study who wore another SenseCam device on the same day. Whatever the case, subjects were asked to arrange the pictures in the order in which they thought they had been taken. Subjects were not given any explicit instructions regarding the origin of the images, they were simply asked to put them in the order in which they thought they had probably been taken.

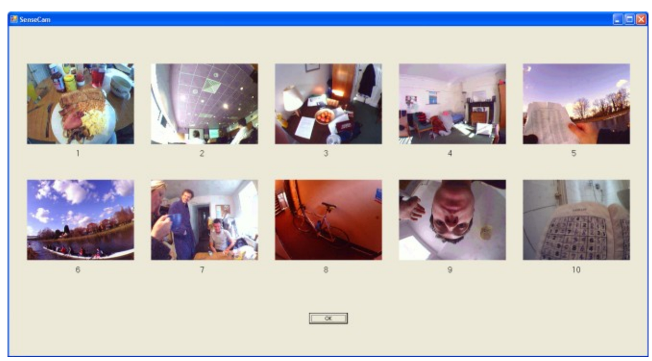

Figure 1. An example of SenseCam images as presented in the temporal "ordering" task.

Final free recall: Once this had been completed, subjects were given another two minutes in which they could add to or amend their account of the first half of that day using a different colored pen. Again, they were asked to tick "Remember", "Know" or "Guess" for these amendments. It was suggested subjects ticked "Know" if something in the images suggested that an event must have occurred, even if it was not truly remembered.

The whole process was repeated for the second half of the same day but using Passive pictures this time if Active had been used before (or vice versa), and their own or the yoked subject's images depending on whether it was a SenseCam or Control day being tested. In total, four such recall tests were carried out per session, covering both the morning and the afternoons of one each of the SenseCam and Control days. Testing was counterbalanced throughout for order of testing for Condition, Mode, and morning or afternoon.

Recognition. After the recall test, subjects were tested to see if they could distinguish their own images from those taken by another person (in this case the same subject they had been yoked to in the recall test). None of the images in this phase had been previously presented.

Subjects were presented with an image and asked to respond "Yes" or "No" on the keyboard as to whether the image was one of their own (irrespective of whether it was taken actively or passively) or not. Subjects were given a maximum of 10 seconds in which to identify the image, making this judgment for 40 randomized images in which half of the images were taken by the participant being tested (ten Active and ten Passive) and half taken by a different participant (ten Active and ten Passive).

Apparatus and Materials

The SenseCam cameras were worn around the neck both for Active and Passive photos (there is no viewfinder) as shown in Figure 2. Images were taken though a fish-eye lens (132 degrees), so that objects at head height (such as faces) were captured. Passive images were taken upon detected changes in infrared radiation and light intensity. In the absence of a detected change, an image was taken every 90 seconds. For more details, see [18].

Figure 2. The SenseCam v2.3 prototype.

Active and Passive images

One of the questions we considered was how to best select the sample of Active and Passive images to use in testing. While subjects were asked to take a minimum of 40 Active images per day, in fact subjects took between 27 and 140 per day. For the Passive images, we had a much larger pool to select from, with an average of 643 captured per day. For every image, a color histogram was generated and a face-detection process performed in order to draw representative samples from both the Active and Passive sets. Images that stood out as different in terms of their color statistics, that had high color entropy, and that contained faces were selected [26]. In this way we generated comparable samples of Passive and Active images in terms of color statistics, appearance of faces and visual clarity.

Follow-Up Test Condition

After preliminary analysis of the data, we began to discover that the difference between the test intervals we chose (3 versus 10 days) may not have been long enough to highlight some of the effects we were beginning to observe. With that in mind, we decided to call back as many of the original subjects as we could, to test them at a much longer retention interval. In this case, we were able to re-test a total of 10 of the original subjects 4 months after first having worn SenseCam. There were, however, some differences in the test procedure, mainly because for many subjects there were not enough new Active images left to test them with. This meant that, while initial recall was carried out in the same way, we had to re-use Active images. We therefore also chose to re-use Passive images so that this would not bias the results. (In fact, they had seen half of each set previously in a recall test, and half in a recognition test). They also did not do the ordering task: rather, subjects were allowed to browse through images as they would when normally using SenseCam software. In the recognition test, we did use new images, but we could only conduct this on Passive images.

We present the results of this "extra-long" interval as much as possible alongside the original results in the analysis noting where we need to be cautious in their interpretation.

RESULTS

Analysis of Recall Results

Recalled events (whether rated as "remember", "know", or "guess") were analyzed by assigning a score of .25 for reporting each of the four elements in an event: the details about "what", "where", "when" and "who". This meant that a score of 1 was assigned to each full event reported. In 95% of cases, when subjects reported any part of an event, they were able to report all the constituent parts. For example, they might report they: "had lunch", "in college", "at 12 o'clock", "with Mark and Sue".

Unfortunately, 4 of the 19 subjects provided incomplete data leaving 15 subjects for analysis. Ten of these same subjects provided data also at the extended 4 month interval.

Analyses of variance (ANOVA) were carried out on data for the 10 subjects tested at all three retention intervals (3 days, 10 days, and 4 months, which we will call the "short", "long" and "extra-long" intervals), and on the 15 subjects tested only at the two shorter intervals. Where appropriate, post-hoc t-tests were performed with a Sidak correction. As we found no substantive differences between the analysis of 10 versus all 15 subjects, we report only the results for the 10 subjects for whom we have a full set of data over the 3 intervals. We also focus only on "remember" and "know" responses.

"Remembered" events prior to viewing images

First let us consider the number of events subjects said they "remembered" (as opposed to "know" and "guess"). Figure

3 shows the mean number of remembered events for each half day recalled *prior* to viewing any images for both SenseCam and Control days, in the Active and Passive conditions, across all three test intervals.

Here we find a clear effect of simply wearing and using the camera. On average, the number of events recalled *before viewing any images* was higher on these SenseCam days than on Control days (F(1,9)=8.02, p<.02).

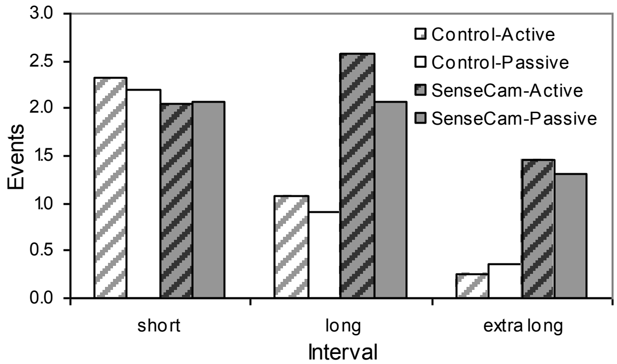

Figure 3. Mean number of remembered events *prior* to viewing images.

The effect of test interval was also significant (F(2,18)=8.42, p<.003). This forgetting effect, however, was dependent on whether these occurred on SenseCam or Control days (as is confirmed by a significant Condition by Interval interaction; F(2,18)= 4.08; p< .03). In other words, as shown in Figure 4, forgetting was more rapid over time for Control days than for SenseCam days, the number of events remembered prior to seeing images dropping to near zero after 4 months.

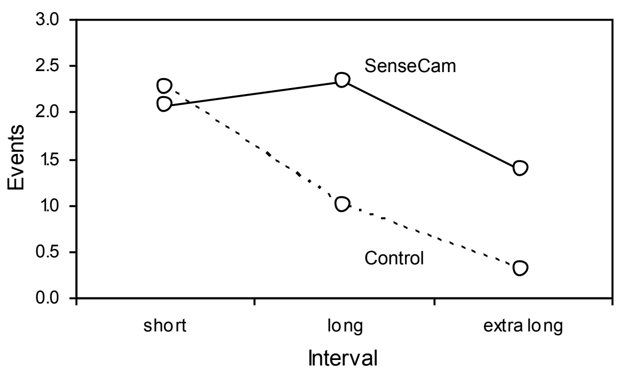

Figure 4. Condition by Interval interaction for number of remembered events *prior* to seeing images.

"Remembered" events after viewing images

The effect of viewing SenseCam images on recall can be examined by looking at the mean number of *additional* events "remembered" after viewing either SenseCam or Control images (Figure 5). Here, we found only a main effect of Condition (F(1,9)=6.42, p< .03). In other words, the data indicate that viewing SenseCam images did indeed give rise to more remembered events than did Control images.

There was no main effect of interval, however this did approach significance (F(2,18)=3.15, p<.07), suggesting an underlying forgetting effect over time. In addition, the lack of an Interval by Condition interaction means that there is no evidence of a differential effect over time of seeing SenseCam versus Control images, as shown in Figure 6.

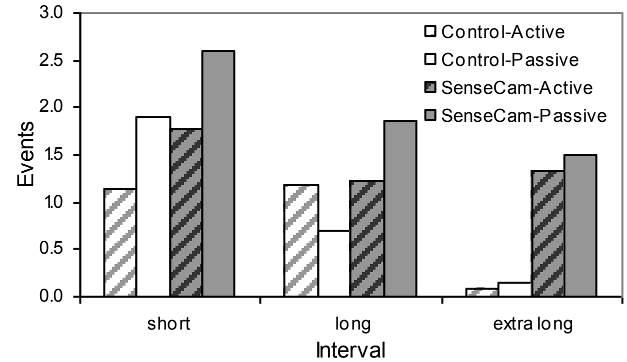

Figure 5. Mean number of additional remembered events *after* viewing images.

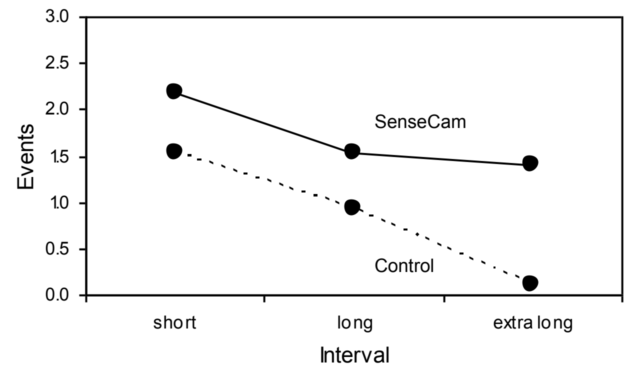

Figure 6. Condition by Interval plot for number of additional remembered events *after* seeing images.

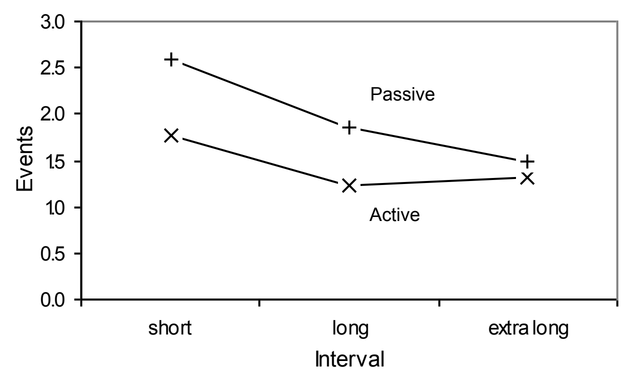

Figure 7. Number of additional remembered events *after* seeing images for Active and Passive images over interval in the Sensecam condition only.

In order to uncover possible differences in Active versus Passive mode of capture, we restricted the analysis only to days when subjects wore SenseCam. Here we found that, contrary to our predictions, Passive images were better triggers for remembered events than Active ones (F(1,9)=6.93, p< .03). This did not depend on test interval (no significant interaction) and there was no main effect of interval (see Figure 7).

"Known" events prior to viewing images

We can now consider events that subjects indicated as ones they were certain that had occurred even though they claimed not to really remember them. On average, subjects reported about 1 "known" event per half day *prior* to viewing images, regardless of whether or not they were wearing SenseCam, how long afterward they were tested, and whether these were Active or Passive test days. There were no main effects or simple interactions.

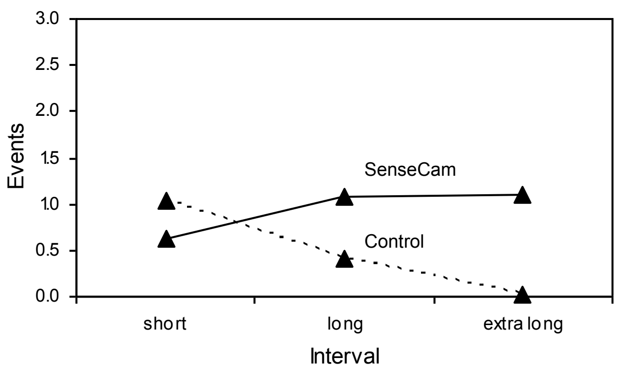

Figure 8. Condition by Interval interaction for number of additional known events *after* seeing images.

"Known" events after viewing images

With regard to the number of additional "known" events *after* viewing images, we find only a significant interaction between Condition and Interval (F(2,18)=7.00, p<.006), as shown in Figure 8. Further, post hoc tests show that while there was a significant decline over time in the Control condition (p<.03), there was no such effect in the SenseCam condition. The number of 'known' events appeared greater for SenseCam versus Control images generally, although this trend only neared significance (F(1,9)=4.02, p<.08). This suggests that the ability to "know" something occurred does not vary with time when viewing SenseCam. However, without images, the ability to know what happened does decline over time, giving rise to a bigger advantage for SenseCam at longer time intervals.

No significant effect of Mode resulted when restricting the analysis to the SenseCam condition only. In other words, unlike "remembered" events, there was no evidence for any advantage for either Active or Passive images in giving rise to "known" events.

Analysis of Ordering Results

One of the drawbacks of the free recall test is that we have no real way of verifying the accuracy of the events that subjects report. The ordering task does not suffer from this problem as we do know the correct order in which the images were captured. The ability of subjects to correctly order images was assessed by calculating a value of Kendall's Tau which allows the comparison of the actual order of images against the order observed to be produced by the subjects. Tau is a correlation coefficient and hence a value near zero indicates chance performance, and values close to 1 or -1 indicate a strong (positive or negative) correspondence between actual and observed orderings.

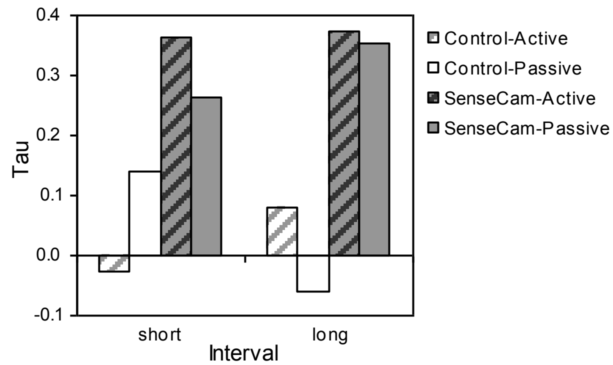

Figure 9. Mean values of tau for the ordering task.

Figure 9 shows the mean Tau values over Condition and Mode of capture, measured at both the short and long time interval (no ordering task was carried out at the extra-long interval). In total, 16 subjects completed this task.

As can be seen, there is a clear effect of whether or not subjects viewed their own as opposed to a control subject's images (borne out a main effect for Condition: F(1,15) = 28.9, p<.001). This is further confirmed by the finding that the ability to put the images in sequence for one's own images is significantly different from zero (F(1,15)=50.62, p<.001). No evidence was found for differences over test interval, between Active and Passive images (even when restricting the analysis to SenseCam days only), and no interactions were found.

Analysis of Recognition Results

Finally, turning to the results of the recognition test, we find a similar pattern of results to the ordering task: subjects were indeed able to effectively distinguish their own images from those of other subjects. While Figure 10 shows that, while subjects occasionally misidentified a Control image as their own, this occurred infrequently. Over 80% of the time, subjects were able to correctly identify their own images. (Note that only Passive images were tested at the extra-long interval.)

ANOVA was conducted on d-prime scores (a standard measure of discriminability, see [16]) for 15 subjects at the two shorter intervals, and 9 subjects at all three intervals. The d-prime scores were found to be significantly different from zero (F(1,14)=552.77, p<0.001, and F(1,8)=316.3, p<0.001) but no other main effects or interactions were found. In other words, there was no evidence that Active images were any more or less recognizable than Passive images, and no evidence that recognition ability declined over time. In this last respect we must be somewhat cautious: At the extra-long interval, although the Passive images used had never been viewed before, they had been tested about events on these days before, so may have been reminded about the events the images depicted.

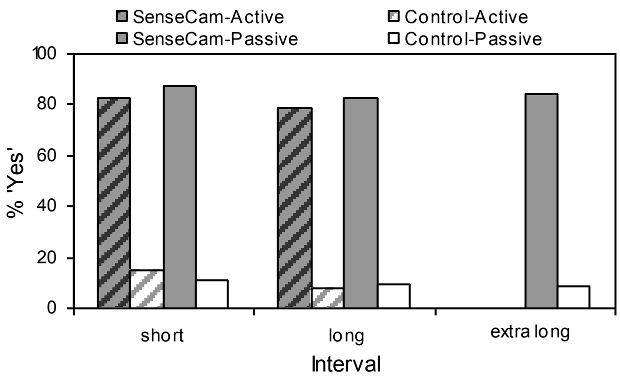

Figure 10. Percentage of time subjects responded "yes" when asked to identify an image as their own in the recognition task.

DISCUSSION

We begin by looking at our first research question: whether SenseCam images significantly improve people's memory for past personal events. The simple answer is "yes" but the data from this study paint a richer picture about *how* such images make connections to one's personal past.

Identifying and Ordering Images

For example, we might surmise that a first test of the meaningfulness of SenseCam images (meaningful on a personal level) is whether people can even recognize such images as being a product of their own activities. In fact we did find that subjects were highly successful at distinguishing their own images from that of Control subjects, and that, furthermore, they were as good at recognizing them after 4 months as they had been at 3 days.

The results further show that, not only could subjects distinguish their own images from those belonging to others, but they were also very good at ordering them. While, for technical reasons, we could not test this ability at a 4 month interval, certainly they were as good at ordering their own images after 3 days as they were on a different set of their own images a week later.

This raises interesting questions about how it is that subjects accomplished both the recognition and ordering tasks. One possibility is that the images triggered a true "remembering" of the associated events in the sense that the subjects could in some way mentally re-experience that event. This then could lead to recognition, or in the case of ordering, help them to place each remembered event alongside other events similarly triggered by other images. Another possibility is that subjects were able to use schematic knowledge about their own routines, people in their lives, familiar places and so on (cf Conway [7]) to identify images as their own, and place them in sequence. A third possibility is that the images themselves contained semantic information which is helpful in ordering images, such as lighting, geographical information (such as an office versus a town centre) or something more specific such as images of people eating lunch. This last possibility would require no personal or autobiographical knowledge, only general semantic knowledge.

The results of the ordering test would lead us to rule out the last conjecture since subjects were only able to order Control images at chance level. Thus we must assume that the images in this study did not contain enough information in and of themselves to allow people to deduce any accurate sort of ordering from them. In other words, we must assume that some kind of autobiographical knowledge is necessary, whether it is of specific events that occurred, or more general schematic knowledge about what one tends to do, where one tends to go, who one knows and so on. Presumably, this autobiographical knowledge is also helping subjects reject images that were not theirs in the recognition test.

Remembering versus Knowing About the Past

The results of the recall test help to explicate the nature of this autobiographical knowledge, and furthermore provide evidence that there are at least two different kinds that SenseCam images support. In the recall test, we made an a priori distinction between events that subjects reported they "remembered" as against those they reported "knowing" must have happened, despite the fact they could not remember them. This is important in relation to the claims that one might make about life-logging technologies. On the one hand, one might claim (and many people do) that captured data will help people to truly remember, reminisce about, and relive past events from one's life. On the other, one may claim that life-logging data will help people know what has occurred, and not be concerned with whether they truly remember the original events in question. The latter may apply to cases where life-logs help people find lost objects, or retrieve details of people's names, conversations, or actions, for example. In these cases we may care only that they retrieve these things or this knowledge successfully.

The recall data show that SenseCam images support both the ability to remember and to know what happened in one's personal past. However, it further suggests that the beneficial effects of these images operate differently in these capacities across time.

Specifically, in the case of remembered events, we failed to find an interaction with the length of the test interval. In other words, there was no evidence that SenseCam images provided any greater (or lesser) benefit over time, at least in the context of the intervals of time we investigated (between 3 days and 4 months). Although we can only conjecture at this point, the evidence suggests that the ability for SenseCam images to trigger true recollection may decline with general forgetting. In other words, over time we naturally forget events; with SenseCam images, we can remember more, but the power of these cues to spark remembering also deteriorates over time. This is suggested by the apparent parallelism of the curves in Figure 6. This conjecture may of course be proven wrong at much longer time intervals, but it is a reasonable inference within the confines of this study.

By way of contrast, we found that the ability to "know" that something occurred after viewing SenseCam images shows greater stability and greater advantage over time. Here we found that subjects were as able to deduce or infer something about their activities from the images at 3 days as they were at 4 months. In other words, the power of SenseCam images to help in this way did not decline over time, and in fact appeared to stay relatively constant. This is all the more striking when we consider that the data also show the ability to "know" something occurred when presented with Control images did significantly decline over time.

Thus we can now address both the second and third research questions by summarizing the above. With regard to the ways in which SenseCam images make connections to people's personal past, we have shown that not only can they recognize their own images through autobiographical knowledge, but that these images can also evoke knowledge of the past in at least two different ways: by helping people to recollect those events, and by helping them to know what happened even in the absence of recollection. With regard to the question of how these abilities change over time, at least within a time window of 4 months, we have seen that the ability to distinguish and order one's own images stays relatively constant. At the same time, whereas the potency of such images in triggering recollection appears to decline with general forgetting, using these images to deduce or know about past events appears less sensitive to the passage of time.

One flaw in the experimental design, but an interesting result nonetheless, is that it is clear that simply wearing and using the SenseCam device helped subjects recall events at the long and extra-long intervals even without viewing any images. It seems that experience of using the device itself gave subjects the necessary "hook" to later pinpoint and remember events occurring on the day in question. In fact, the interviews at the end of the extra-long test session confirm this. Many subjects reported on occasions in which people commented on their wearing the device, or could remember interesting or funny details of occasions in which they took pictures. We must therefore take this effect into account in interpreting the recall data. While our analysis considers additional events reported *over and above* this baseline recall level (after seeing images), it may be that the higher baseline level for SenseCam days raises the possibility that other related events will be recalled *after* seeing images. In a future study, it would be advisable to control for the "novelty effect" of wearing SenseCam by having subjects wear it even on Control days.

Active versus Passive Images

As a final point of discussion we turn to the interesting issue of the Active versus Passive capture. We surmised in advance that Active images would be more personally meaningful than Passive images in that subjects could select which events to depict, and could, to some extent, control the kind of image that was taken. This led us to predict not only that Active images would be easier to recognize and put in sequential order, but that they would also act as more powerful cues for recall.

In fact we failed to find any difference in people's ability to recognize or order Active and Passive images. As long as the images were theirs, they could do both tasks equally well with either kind. This suggests that the kind of Passive images provided by SenseCam contain just as much autobiographically relevant material as Active ones. This may be attributable to the success of the design of SenseCam in its use of a fish-eye lens, or in the algorithm it uses for capture. Alternatively, it may be that the lack of a viewfinder makes it difficult to adequately control and frame the kinds of Active images SenseCam can take.

A further surprise, and a finding counter to our prediction, was that the Passive images were found to be *better* cues for recall than the Active images. (Note, this is contrary to the "generation effect" [28], discussed earlier.) Why this is the case we can only conjecture about. One possibility is that the act of taking pictures somehow interferes with forming a memory of that event, although the fact that our subjects tended to remember and tell us about particular occasions on which they took pictures would suggest this is not the case. Another possibility is that the kinds of events that people tend to remember anyway – the most salient ones—are the events during which Active images are taken. Thus, seeing these images may be somewhat redundant. This is testable in future work by looking carefully at the connection between remembered events and the events depicted in the images.

CONCLUSIONS

This study provides evidence that at least some kinds of cues captured by life-logging technologies, in this case SenseCam images, can be shown to provide effective links to events in people's personal past. Further, it suggests that the automatic way in which SenseCam captures these images results in cues which are as effective in triggering memory as images which people capture on their own initiative. In fact, with regard to the recollection of past events, these passively captured images may even cause people to remember *more* events than they would with their own actively-captured images.

But beyond the particular results here, the study highlights the fact that we need to be clear about the claims we are making about the benefits of life-logging technologies. This includes thinking about what value such systems will have for people in the long term. For example, this study raises the possibility that the potency of images as cues for remembering might not be effective in the very long term. On the other hand, such cues may well retain the ability to help us *know* about our past over time. Further work looking at much longer time intervals needs to be carried out to substantiate these claims. We also need to explore whether other kinds of cues (or combinations of them) would help us support people more effectively in the ways we would like. Rather than to speculate or rely solely on our own best guesses, we suggest that systematic experiments such as this can help us make more informed decisions based on a deeper understanding of the relationship between life-logging technology and human memory.

ACKNOWLEDGMENTS

Many thanks go to Tom Powell who helped to run and analyze the study, to James Srinivasan who provided technical assistance with SenseCam, and to Lyndsay Williams, the inventor of SenseCam. We are also grateful to Antonio Criminisi for help with the algorithm used in selecting test images, Shahram Izadi, Gavin Smyth and Tim Regan for programming help, and Emma Berry, Georgina Browne and Andrew Fitzgibbon for comments on the paper.

REFERENCES

1. Adar, E., Karger, D.R., and Stein, L. (1999). Haystack: Per-user information environments. *Proc. Information and Knowledge Management.*
2. Bell, G. (2001). A personal digital store. *Comms of the ACM, 44(1), 86-91.*
3. Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B. and Wood, K. (2006). The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis. *Neuropsychological Rehabilitation*.
4. Brewer, W. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering Reconsidered*. New York: Cambridge University Press, 21-90.
5. Bush, V. (1945). As we may think. *Atlantic Monthly*.
6. Carter, S. and Mankoff, J. (2005). When participants do the capturing: The role of media in diary studies. *Proc. CHI 2005*, 899-908.
7. Conway, M. (1990). *Autobiographical memory*. Milton Keynes: Open University Press.
8. Czerwinski, M. & Horvitz, E. (2002). Memory for daily computing events. *Proc. HCI 2002*, 230-245.
9. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin R. & Robbins D. (2003). Stuff I've Seen: A system for personal information retrieval and re-use. *Proc. SIGIR 2003.*
10. Eagle, N. & Pentland, A. (2005). Reality mining: Sensing complex social systems. *J. of Personal and Ubiquitous Computing*.
11. Eldridge, M., Lamming, M. & Flynn, M. (1991*). Does a video diary help recall?* EuroPARC Technical Rep. No. EPC-1991-124.
12. Fitzgibbon, A. and Reiter, E. (2004). *Memories for life: Managing information over a human lifetime*. In T. Hoare & R. Milner (Eds.), *Grand Challenges in Computing Research*, (13-16). Swindon: British Computing Society.
13. Freeman, E. & Gelernter, D. (1996). Lifestreams: A storage model for personal data. *ACM SIGMOD Bulletin*.
14. Gardiner, J.M. & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. Craik (Eds.), *The Oxford Handbook of Memory*, (229-244). Oxford: Oxford University Press.
15. Gemmell, J., Bell, G. & Lueder, R. (2006). MyLifeBits: A personal database for everything, *Comms. ACM, 49(1)*, 88-95.
16. Green, D. & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.
17. Hori T. & Aizawa K. (2003). Context-based video retrieval system for the life-log applications. *ACM Workshop on Multimedia Information Retrieval,* 31-38.
18. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., and Wood, K. (2006). SenseCam: A retrospective memory aid. *Proc. Ubicomp 2006*.
19. La Marca, A., Hightower, J., Smith, I., Scott, S., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello G., & Schilit, B. (2005). Place Lab: Device positioning using radio beacons in the wild. *Proc. Pervasive 2005*.
20. Lamming, M., Brown, P., Carter, K., Eldridge, M., Flynn, M., Louie, G., Robinson, P., & Sellen, A. (1994). The design of a human memory prosthesis. *Computer Journal, 37(3),* 153-163.
21. Lifeblog (Nokia). http://www.nokia.co.uk/nokia/0,,71739,00.html
22. Lifelog (DARPA). http://www.darpa.mil/ipto/programs/lifelog/index.htm
23. Linton, M. (1978). Real world memory after six years. In M. Gruneberg, P. Morris, & R. Sykes (Eds), *Practical Aspects of Memory*. London: Academic Press.
24. Mann, M., Fung, J., Aimone, C., Sehgal, A. & Chen, D. (2005). Designing EyeTap digital eyeglasses for continuous lifelong capture and sharing of personal experiences. *Alt.Chi, Proc. CHI 2005*.
25. PlaceLab (MIT). http://architecture.mit.edu/house_n/placelab.html
26. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A., (2006). AutoCollage. *Procs. SIGGRAPH 2006*.
27. Schilit, B., Adams, I., & Want, R. (1994). Context-aware computing applications. *Proc. of the Workshop on Mobile Computing Systems and Applications*, 85-90.
28. Slamecka, N. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn. Mem.,* 4, 592-604.
29. Tulving, E. & Donaldson, W. (1972). *Organization of memory.* New York: Academic Press.
30. Vermuri, S., Schmandt, C., Bender, W., Tellex, S., & Lassey, B. (2004). An audio-based personal memory aid. *Proc. Ubicomp 2004*, 400-417.
31. Wagenaar, W. (1986). My memory: A study of autobiographical memory after six years. *Cognitive Psychology, 18,* 225-252.
32. Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The Active Badge location system. *ACM Trans.on Information Systems*, *10(1),* 91-102.