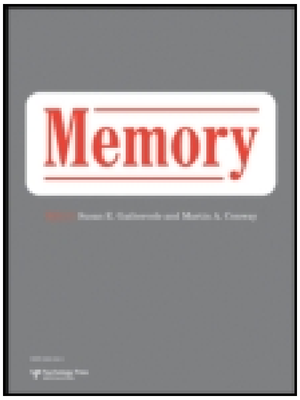

Memory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam

Jason R. Finley [a], William F. Brewer [a] & Aaron S. Benjamin [a]
[a] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Available online: 24 May 2011

MEMORY, 2011, 19 (7), 796–807

Psychology Press
Taylor & Francis Group

The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam

Jason R. Finley, William F. Brewer, and Aaron S. Benjamin

Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Emerging ‘‘life-logging’’ technologies have tremendous potential to augment human autobiographical memory by recording and processing vast amounts of information from an individual’s experiences. In this experiment undergraduate participants wore a SenseCam, a small, sensor-equipped digital camera, as they went about their normal daily activities for five consecutive days. Pictures were captured either at fixed intervals or as triggered by SenseCam’s sensors. On two of five nights, participants watched an end-of-day review of a random subset of pictures captured that day. Participants were tested with a variety of memory measures at intervals of 1, 3, and 8 weeks. The most fruitful of six measures were recognition rating (on a 1–7 scale) and picture-cued recall length. On these tests, end-of-day review enhanced performance relative to no review, while pictures triggered by SenseCam’s sensors showed little difference in performance compared to those taken at fixed time intervals. We discuss the promise of SenseCam as a tool for research and for improving autobiographical memory.

Keywords: Autobiographical memory; Experiences; Events; Life-logging; SenseCam; Photographs.

The study of human cognition has much to gain from advances in technology. Increasingly compact and powerful devices—such as SenseCam, a wearable, sensor-equipped digital camera—enable the automated chronicling, or ‘‘life-logging’’, of many aspects of a person’s daily experience (see Bell & Gemmell, 2007, 2009). One research domain in which such technology has great potential is the study of autobiographical memory: ‘‘memory for information related to the self’’ (Brewer, 1986, p. 26). Such research explores the ways in which humans maintain and remember personal experiences, and how those memories serve to ‘‘support a coherent and effective self’’ (Conway, 2005, p. 596).

Serious methodological challenges are posed by the study of autobiographical memory, as people’s everyday experiences are not experimentally controlled, can span decades, and are largely unknowable to others (Brewer, 1986, 1996; Conway, 1990). Galton (1879a, 1879b) created a technique for the study of autobiographical memory (later modified by Crovitz & Schiffman, 1974) that used experimenter-chosen probe words to elicit participants’ associated memories. This method suffers from potentially unrepresentative sampling of memory and inconsistency in retrieval and response strategies across participants. Another approach to the study of

Address correspondence to: Jason R. Finley, Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, Illinois 61820, USA. E-mail: jrfinley@illinois.edu

This research was supported by a grant from Microsoft Research, Ltd., Cambridge, UK (grant # 2007-066), and from the National Institute of Health to ASB (grant # R01 AG026263). We thank Ellen Brewer for comments on an earlier version of this manuscript. Special thanks to James Scott, Gavin Smyth, Alex Butler, and Steve Hodges for superb technical support.

http://www.psypress.com/memory http://dx.doi.org/10.1080/09658211.2010.532807

autobiographical memory used the contents of long-running daily diaries to cue and assess memory (Linton, 1975; Wagenaar, 1986), but this method suffers from biased selection of experiences by the participants themselves. Advances in technology enabled the advent of experience sampling methods (Conner, Barrett, Tugade, & Tennen, 2007), such as Brewer's (1988) experiments in which participants carried small electronic devices that prompted them at random intervals to record certain aspects of their current experience on paper response cards. Portions of these cards were later used to test memory. This method suffers the disadvantage of frequently and actively involving participants in the capture of information, potentially altering the very memories it sets out to measure. SenseCam can unobtrusively capture a pictorial record from the lives of individuals, thus opening a new route to the investigation of autobiographical memory.

SenseCam also has potential for augmenting human cognition: as a cognitive prosthetic for people with memory impairments (Berry et al., 2007) or a memory aid for those with unimpaired memory (Sellen, Fogg, Hodges, & Wood, 2007). In this paper we consider how the capabilities of the device and the nature of its use can serve to investigate and enhance autobiographical memory. Specifically, we evaluate whether end-of-day review of SenseCam pictures enhances memory for the day's events, and also whether the sensors built into SenseCam capture more memorable pictures than those captured using a simple interval timer.One way in which SenseCam can augment memory is by providing pictures with which one can easily review a day's experiences. It has long been known that repeated presentation of stimuli improves memory (Ebbinghaus, 1885/1913). More specifically, prior work has shown that reviewing still pictures from a previously viewed movie of everyday events enhances performance on subsequent free recall and recognition (Koutstaal, Schacter, Johnson, Angell, & Gross, 1998; Schacter, Koutstaal, Johnson, Gross, & Angell, 1997). Koutstaal, Schacter, Johnson, and Galluccio (1999) found similar effects on free recall (Experiment 1) and recognition combined with cued recall (Experiment 2) when participants performed a variety of activities in the laboratory and then reviewed photographs of actors performing those same activities. The current experiment sought to extend these findings to the investigation of unconstrained personal activities, review of first-person pictures, and longer retention intervals. We predicted that an end-of-day review of SenseCam pictures would improve memory.

Life-logging technologies like SenseCam can produce a wealth of durable external personal memory; however, this raises new problems of information management, organisation, and searching (Jones & Ross, 2007). One way to ameliorate these problems is to strategically filter which data are captured. SenseCam can do so by using sensors (white light, passive infrared, and accelerometer) to capture pictures potentially relevant for future memory needs. If there is a systematic relationship between objectively measurable physical characteristics of experience (e.g., incidence of white light) and psychological characteristics of experience (e.g., memory for an event), and if SenseCam's sensors and triggering algorithm are well calibrated to this relationship, then we would expect sensor-triggered pictures to elicit memory better than pictures simply captured at regular intervals throughout the day.

The goal of the present experiment was to gather data that would allow us (a) to evaluate a variety of potential memory measures, (b) to compare memory for days that had been reviewed with memory for days that had not been reviewed, (c) to compare memory for a set of pictures triggered by SenseCam's sensors with a set taken at regular time intervals, and (d) to examine the course of forgetting under each of these conditions.

METHOD

Participants

There were 12 undergraduates who participated for payment of $50 each. There were 4 males and 8 females, ages 18 to 21.

Recruitment and selection

Participants were recruited via posters and web notices. Of the 24 undergraduates interviewed and who returned completed questionnaires, 12 participants were selected on the basis of their punctuality, ability to follow directions precisely, self-rated responsibility in caring for electronic devices, and having living situations and daily activities that would least interfere with wearing a SenseCam for a considerable portion of each day.

Privacy

Several aspects of the experimental procedure were designed to safeguard the rights and privacy of participants and others. Potential participants were told to complete the initial questionnaire only if they were comfortable with the full procedure as described in the interview. Each participant's data were identified only by a unique code number assigned to each individual. Pictures from an individual participant's acquisition period were never seen by anyone except the experimenters and that participant.

Participants could deactivate the SenseCam (using the on/off button or the privacy button) at any time that they chose to avoid the possibility of pictures being captured. Pressing the privacy button deactivates picture capture for 7 minutes, with a warning beep sounding 15 seconds before reactivation. Participants were further instructed to deactivate the SenseCam in several specific situations: restroom, changing/locker room, doctor's office, ATM or bank, and any time someone else requested that it be deactivated. Furthermore, when entering their workplace or a private residence, participants were instructed to deactivate the camera until they received permission to reactivate it. Finally, participants could note time periods for deletion in the small notepad they carried (e.g., if they suddenly realised they had left the SenseCam on while using the restroom). All pictures falling in such time periods were deleted at the end of each day without having been seen by anyone.

Apparatus

SenseCam is a small, wearable, sensor-equipped digital camera developed by the Microsoft Research lab in Cambridge, UK. The SenseCams used in this experiment were hardware version 2.3b and firmware version 2.6.7. A built-in adjustable lanyard allows SenseCam to be worn around the neck at upper chest height. Pictures are captured using a fisheye lens with a wide angle of view (119° diagonal) and stored with timestamps on a 1Gb SD flash memory card. The pictures are saved as JPEG files with display size 640×480 pixels, resolution 72 dots per inch, colour depth 32-bit (16.7 million colours), and file size approximately 30 kilobytes. The maximum rate of picture capture is approximately once every 5 seconds. Capture may be triggered by three different methods, each of which can be independently enabled or disabled by experimenters: (a) manually (by pressing the manual trigger button); (b) by timer; (c) by sensors. The device gives no audible or visible indication of picture capture due to timer or sensor triggering, although it beeps upon manual triggering.

SenseCam sensors

SenseCam has sensors for measuring: (a) temperature (in units of .5° C); (b) white light intensity (on a scale of 0 to 10,000); (c) change in infrared radiation (e.g., from a moving warm body) as either presently occurring (1) or not (0) within a range of approximately 2.5 m for a human-sized object; and (d) acceleration in three dimensions (in units of 1/819th g). Data from these sensors are recorded, along with date and time, approximately every 1.5 seconds. Pictures and sensor data are transferred to a computer via a USB interface, which is also used for recharging the battery.

Sensor-triggering algorithm

The default algorithm and threshold values for sensor triggering of picture capture were established by Microsoft Research through informal testing. The algorithm compares current sensor values to the set of values most recently measured (typically about 1.5 seconds prior). If a picture is not already currently being captured and saved, a new capture will be triggered if either (a) there is a change in white light level greater than 4500 units as measured by the detector, or (b) the change in accelerometer value on all three axes is no greater than 500 units as measured by the detector (indicating that the device is probably stationary) *and* the passive infrared value is 1 (indicating change in infrared, typically due to nearby human movement). More detailed specifications can be found in Hodges et al. (2006).

Design

The experiment used a $3 \times 2 \times 2$ within-participants factorial design with three retention intervals (1, 3, 8 weeks); two types of trigger condition (sensor-triggered, timer-triggered); and two types of review condition (no review, review). The order in which participants received each trigger and review condition was counterbalanced using a balanced Latin Square.

Acquisition period

Participants wore a SenseCam as they went about their normal daily activities for five consecutive weekdays (i.e., Monday through Friday). The first day was treated as practice, and pictures from that day were not reviewed or tested. SenseCam operated automatically throughout the day, capturing pictures according to the trigger condition for that day (described below).

The starting and ending time of acquisition for each day were determined by each individual participant, and thus varied by participant and by day. For the four experimental days, the mean starting time was 09:55 ($SD = 1$ hr 11 min), the mean ending time was 19:17 ($SD = 1$ hr), and the mean duration of acquisition per day was 9 hours and 22 min ($SD = 1$ hr 37 min, range = 5 hr 13 min to 12 hr 43 min).

Participants were informed before the acquisition period that their memory for that period would later be tested, and were told the general nature of the questions they would be asked (e.g., ''You may be asked to write down details about the events that occurred during the days on which you wore the SenseCam, either cued by images from those days or with no such cues'').

At the end of each day participants returned to the lab. The experimenter transferred all of the day's pictures from the SenseCam onto a computer, deleted any pictures falling in a time period marked for deletion by the participants in their notepads, and updated the SenseCam's trigger setting for the following day. Participants then received a review of that day's pictures (described in detail below), if the day had been assigned to be a review day. This entire procedure was repeated for 5 days.

Trigger condition. Manual triggering of picture capture was disabled for the experiment. On any given day, the SenseCam captured pictures either as triggered by its sensors (see Apparatus), or as triggered by its fixed-interval timer. The measured mean timer interval for the timer-triggered condition was 10.6 seconds ($SD = 1.3$). Note that because timer-triggered pictures were not triggered by any external criteria, and because pictures were randomly selected for review and for testing, this condition provides a de facto random sampling of experience and thus serves as a logical control condition against which to test sensor triggering.

End-of-day review. The end-of-day review was a movie composed of a sequence of randomly selected pictures captured by the SenseCam on that day for that participant. The review was shown on a computer monitor with pictures presented one-at-a-time for 1 second each in chronological order. Participants were told that they would be watching a movie composed of a sequence of pictures taken throughout that day; they were given no special instructions on how to process the review. Each participant received two end-of-day reviews, one on a sensor-triggered day and one on a timer-triggered day.

The number of pictures used to create the review movie was half of the number of pictures captured on the practice day for that participant (which was always the first day and used sensor-mode without review). We could not use the same number of pictures for all participants because it was not known ahead of time how many pictures would be captured for each participant on each day. Our main concern was to ensure that, for each individual participant, the same number of images was used for both end-of-day reviews. A participant's first day of SenseCam use provided an estimate of the number of daily pictures we could expect to be captured for that individual. To account for the possibility of fewer pictures being captured on subsequent days, we conservatively set the number of pictures to be used for review as half of the number captured on each participant's first day ($M = 1{,}022$, $SD = 235$). The mean duration of the reviews was 17 min 2 s ($SD = 3$ min 55 s).

Memory measures

Memory for events that occurred during the acquisition period was tested with six different test procedures:

Recollection Judgement. Participants were shown a picture and made a *yes/no* judgement in response to the following instructions: ''Recollective memory is when you are able to recall a specific episode or moment from your life. It is almost like reliving or re-experiencing the past event. Please indicate whether you have a recollective memory for this picture.''

Recognition Rating. Participants were shown a picture accompanied by the following instructions: ''Please rate the strength of your memory for the scene(s) and/or event(s) depicted by this picture.'' Participants responded using a 1–7 scale, with 1 labelled as ''no memory'' and 7 labelled as ''extremely strong memory''. In contrast to the standard old–new recognition method used in laboratory experiments, no lures were presented. Since lures could be made arbitrarily easy or difficult to correctly reject (e.g., a picture taken on the Moon vs a picture taken from 10 cm to the participant's right), we saw no value in including them.

Picture-Cued Recall. Participants were shown a picture accompanied by the following instructions: ''Please describe in detail the visual scene(s)/event(s) as they would appear to the SenseCam in the one minute following this picture. If you can't remember anything for this picture, you may type a question mark to continue to the next picture.'' There was a maximum of 1 minute to respond to each item, with no minimum time requirement.

Time-Cued Recall. Participants were shown a date and time from their acquisition period (e.g., ''Tuesday 11:09 am'') accompanied by the following instructions: ''Please describe in detail the visual scene(s)/event(s) as they appeared to the SenseCam at the above day and time. If you can't remember anything for this day and time, you may type a question mark to continue to the next trial.'' There was a maximum of 1 minute to respond to each item, with no minimum time requirement.

Recall of Time. Participants were shown a picture and asked to give the day and time it was taken. They made a forced choice on *day* from among the days they wore SenseCam (e.g., Monday, Tuesday, Wednesday, Thursday, Friday), and a forced choice on *time of day* (12:00 am to 11:55 pm in intervals of 5 minutes). A response was required for each field, and participants were instructed to give their best guess if they were not sure.

Temporal Order Judgement. Participants were shown two pictures arranged vertically on the screen and they made a forced-choice judgement about which was captured first. Paired pictures were selected from the same acquisition day (and thus from the same review condition), with acquisition order counterbalanced.

Memory test procedures

Tests were administered individually on computers with instructions presented entirely via onscreen text. Pictures were selected randomly (without replacement) from each participant's total set of captured pictures for inclusion on the tests for that participant. Pictures were never tested more than once. For a given test session each participant completed the following number of items: Recollection Judgement (192); Recognition Rating (192); Picture-Cued Recall (48); Time-Cued Recall (8); Recall of Time (96); Temporal Order Judgement (192). Fewer trials were given for the more time-consuming memory measures in order to reduce the total duration of the testing sessions.

Counterbalancing ensured equal representation in each phase of each test for each participant for all manipulated variables, as well as whether or not the picture had been included in the review if it came from a review day. Participants were only shown and tested on pictures from their own acquisition periods. They had been informed about this aspect of the experiment, so they knew that all of the pictures were ones from their lives.

Each test consisted of four phases, with participants allowed to take 5-minute breaks between phases. Each of the first three phases included five blocks, with each block consisting of only one type of item (all but the Time-Cued Recall items). Block order was determined randomly. The fourth phase consisted of the Time-Cued Recall items.

Memory test intervals

Participants returned to the lab for testing sessions at retention intervals of approximately 1, 3, and 8 weeks. The mean number of days that the testing sessions occurred after the final acquisition day are as follows: Test 1: 5.67 days ($SD = 2.01$, $range = 4–10$); Test 2: 29.50 days ($SD = 1.12$, $range = 27–31$); Test 3: 57.50 days ($SD = 2.43$, $range = 52–62$).

Payment schedule

Participants were each paid $50 in cash in four instalments: $15 at the end of the acquisition period, $10 upon arriving for the first and the second test sessions, and $15 upon arriving for the third (final) test session.

RESULTS AND DISCUSSION

For all tests reported below, an alpha level of .05 was used to indicate statistical significance. Effect sizes for ANOVAs are reported as estimated partial ω^2 calculated using the formulae provided by Maxwell and Delaney (2004).

Pictures captured

For the four experimental days, the mean number of pictures captured per day was 2390 ($SD = 566$, $range = 1358–3479$) in the timer-triggered condition and 2031 ($SD = 565$, $range = 721–3200$) in the sensor-triggered condition. The mean number of pictures captured per hour was reliably greater for the timer-triggered condition ($M = 272$, $SD = 60$) versus the sensor-triggered condition ($M = 209$, $SD = 60$), $t(11) = 3.37$.

Of pictures captured in the sensor-triggered condition, 95% were triggered by passive infrared (which can only be triggered when the camera is stationary), 4% were triggered by white light level, and 1% were triggered by both. Thus, for the types of events in the weekday lives of American undergraduates, the sensor algorithm was most often triggered by the passive infrared sensor. This may reflect the sedentary nature of common activities in undergraduates' daytime weekday lives (e.g., sitting in classrooms, working at computers, watching television) along with their interactions with sources of infrared (e.g., motion of other people or of the participant's hands in front of the device).

Memory measures: Scoring criteria

The six memory measures were scored as follows: (a) Recollection Judgements: 0 for "no" and 1 for "yes," (b) Recognition Rating: 1 to 7, where 7 was "extremely strong memory," (c) Picture-Cued Recall and Time-Cued Recall: amount recalled as response length in number of characters (length = 0 for "?" responses), (d) Recall of Time: 0 to 2, with one point each for correct day and correct time to within one hour, and (e) Temporal Order Judgement: 0 for incorrect and 1 for correct.

Overall data

In the results and discussion section we focus on those aspects of the data that let us answer the questions we set out to study. However, for completeness we give the entire data set for all six memory measures as a function of retention interval, review condition, and trigger condition in the Appendix.

Evaluating the memory measures

We assumed that over 8 weeks there would be substantial forgetting, enabling a rough comparison of the sensitivity of our various memory measures. We reasoned that the greater the decline in performance across tests for a particular measure, the greater the degree to which actual retrieval of episodic details was required for that measure, and thus the more sensitive it is as a measure of autobiographical memory. The values for each memory measure were first standardised within participants, to account for individual differences in absolute levels of performance and to allow direct comparison of memory measures. The mean of the participant means of these standardised values, across retention interval, is shown in Table 1. The right-most column shows the difference from the first test to the third test.

The *Picture-Cued Recall* measure showed the largest amount of forgetting. One would expect a standard probed recall measure for the content of experienced events to be a good response measure. It appears to work well using just the amount written down without scoring for accuracy. The *Recognition Rating (1–7 scale)* for the pictures was the next most sensitive measure. This is consistent with previous studies of autobiographical memory that found this type of measure to be effective (e.g., Brewer, 1988). The data for the *Time-Cued Recall* task suggest that time was not as good a probe for recall as were the pictures of the actual events. This is somewhat analogous to the findings reported in Brewer (1988) that time was a relatively poor

TABLE 1
Memory measures: Means of participant means of scores standardised within participants

	Retention interval			
Memory measure	*1 week*	*3 weeks*	*8 weeks*	*1 week − 8 weeks*
Picture-Cued Recall	.37	−.10	−.28	.65
Recognition Rating	.29	−.07	−.22	.50
Time-Cued Recall	.30	−.10	−.20	.49
Recollection Judgement	.18	−.07	−.11	.29
Temporal Order Judgement	.02	.03	−.05	.06
Recall of Time	.00	.05	−.05	.05

probe cue in autobiographical memory. Furthermore, we think that when given a time and a day of the week, participants could have used knowledge of their weekly schedules during that semester to infer where they would have been and what the visual scene would have been, without necessarily having any specific episodic memory for that time. The simple binary *Recollection Judgement* shows some modest memory sensitivity. Note that in some ways this measure can be thought of as a binary version of the more sensitive 1–7 recognition scale. The *Temporal Order Judgement* and *Recall of Time* measure show little sensitivity to forgetting. Friedman (1993) has made powerful arguments that much time information in autobiographical memory is reconstructed from knowledge of script-based information. We think it is likely that these two measures reflect this type of reconstruction and thus do not show episodic forgetting over the time intervals studied. For example, it would have been trivial for participants to correctly order a daytime and a night-time picture from the same day, without any recourse to episodic memory. Similarly, the approximate time of day could be inferred from information in the picture along with participants' knowledge of their schedules.

On the basis of the above analysis, and for the sake of brevity, we have chosen to focus on Recognition Rating and Picture-Cued Recall as the memory measures for the rest of the analyses. They are the top two measures in terms of forgetting effect size, and the use of two different measures (one recognition and one recall) allows us to see how consistent our findings are across different indexes of memory.

Overall ANOVA for Recognition Ratings

A three-way ANOVA was conducted to evaluate the effect of retention interval (three levels), review condition (two levels), and trigger condition (two levels) on Recognition Rating. Mauchly's test indicated that the assumption of sphericity was met for all main effects and interactions. There was a statistically significant main effect of retention interval, $F(2, 22) = 11.19$, $MSE = 1.07$, $\omega^2 = .170$, and of review condition, $F(1, 11) = 5.02$, $MSE = 0.65$, $\omega^2 = .028$, but not of trigger condition, $F(1, 11) = 0.54$, $MSE = 0.69$, $p = .476$, $\omega^2 = -.004$. No interactions were statistically significant. Follow-up analyses are presented in separate sections below.

Overall ANOVA for Picture-Cued Recall

A three-way ANOVA was conducted to evaluate the effect of retention interval (three levels), review condition (two levels), and trigger condition (two levels) on Picture-Cued Recall. Mauchly's test indicated that the assumption of sphericity was violated for the interaction of retention interval with trigger condition, $\chi^2(2) = 8.08$, $p = .018$, and nearly violated for the main effect of retention interval, $\chi^2(2) = 5.41$, $p = .067$. Analyses were conducted using Greenhouse-Geisser adjusted degrees of freedom for these two effects, but because the results did not differ importantly from analyses assuming sphericity, we present the latter here for the sake of consistency. There was a statistically significant main effect of retention interval, $F(2, 22) = 7.09$, $MSE = 1{,}017.54$, $\omega^2 = .129$, and of review condition, $F(1, 11) = 5.24$, $MSE = 175.18$, $\omega^2 = .012$, but not of trigger condition, $F(1, 11) = 3.62$, $MSE = 324.45$, $p = .083$, $\omega^2 = .013$. No interactions were statistically significant.

Forgetting function

Figure 1 shows the decline in memory performance across the three time intervals for both Recognition Rating and Picture-Cued Recall.

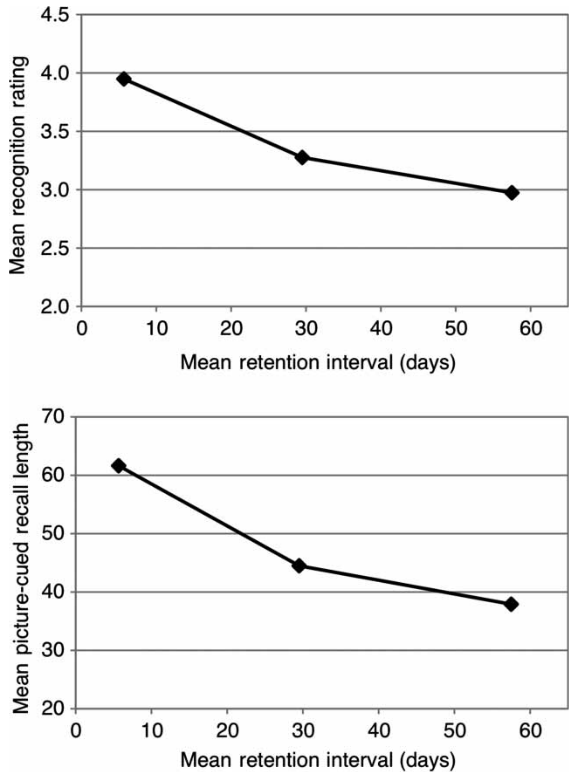

Figure 1. Mean recognition rating (top) and mean picture-cued recall length (bottom) as a function of mean retention interval for the three testing sessions.

Mean performance on both measures decreased reliably across the three retention intervals, as confirmed by separate least-square fittings of a power function ($P = at^b$, where P is performance and t is time) for each participant for Recognition Rating (mean standardized $b = -0.22$, $t(11) = -4.12$, one-tailed), and for Picture-Cued Recall (mean standardized $b = -0.19$, $t(11) = -2.44$, one-tailed). One-tailed tests were used based on the strong theoretical prediction that memory declines over time.

End-of-day review

Figure 2 shows that an end-of-day review improved performance for both Recognition Rating and Picture-Cued Recall. Recognition Ratings were reliably higher for reviewed days ($M = 3.55$, $SD = 0.69$) versus non-reviewed days ($M = 3.25$, $SD = 0.80$). Picture-Cued Recall lengths were reliably higher when there had been an end-of-day review ($M = 50.55$, $SD = 20.85$) versus no such review ($M = 45.51$, $SD = 18.49$).

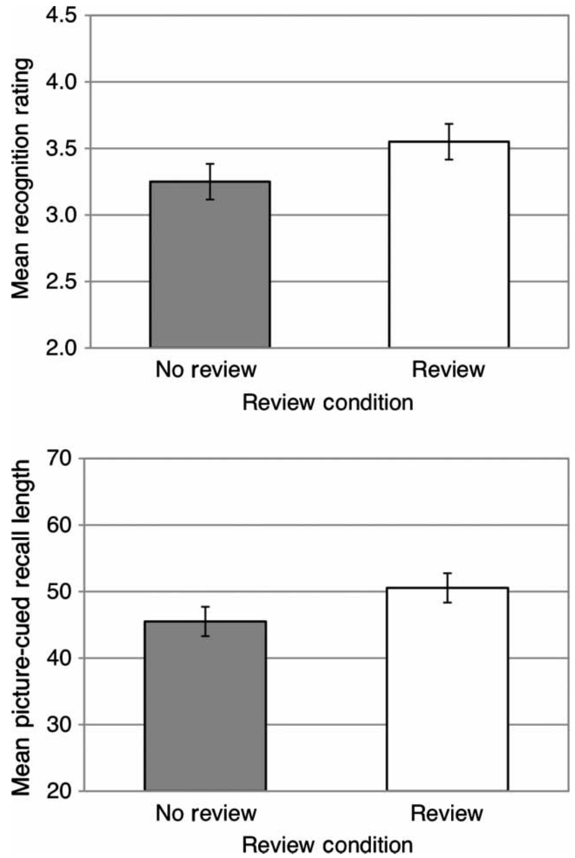

Figure 2. Mean recognition rating (top) and mean picture-cued recall length (bottom) as a function of review condition. Error bars represent standard error of the difference between conditions.

Trigger condition

Figure 3 shows that the default SenseCam sensor settings provide little improvement in selecting items yielding higher memory performance compared to randomly selected items for both Recognition Rating and Picture-Cued Recall. Recognition Ratings were not reliably different for pictures captured as triggered by the sensors ($M = 3.45$, $SD = 0.75$) versus the timer ($M = 3.35$, $SD = 0.75$). Picture-Cued Recall lengths were not reliably different for pictures captured as triggered by the sensors ($M = 50.89$, $SD = 20.05$) versus the timer ($M = 45.17$, $SD = 19.94$). Recognition Ratings were reliably greater for pictures triggered by the passive infrared sensor ($M = 3.47$, $SD = 0.76$) versus the white light level sensor ($M = 2.92$, $SD = 1.13$), $t(11) = 2.83$. There were not enough data to conduct the same comparison for Picture-Cued Recall.

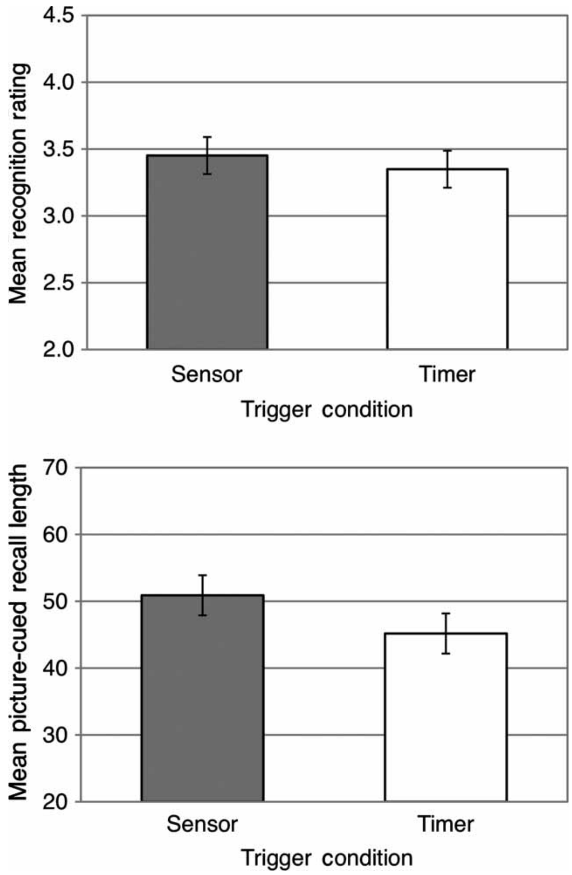

Figure 3. Mean recognition rating (top) and mean picture-cued recall length (bottom) as a function of trigger condition. Error bars represent standard error of the difference between conditions.

GENERAL DISCUSSION

The results showed substantial forgetting across retention intervals, and also revealed that certain memory measures were better indicators of episodic memory than others (i.e., they were more sensitive to forgetting), as shown in Table 1. Performance was greater for reviewed days than for non-reviewed days, but did not reliably differ for sensor-triggered versus timer-triggered pictures.

Our data suggest that an end-of-day review of an individual's activities (via pictures captured by SenseCam) increases performance on several memory measures for that day. However, we need a deeper understanding of the factors involved in the review process. Such factors include (a) types of pictures chosen for review, (b) presentation order (e.g., non-chronological ordering might actually enhance memory), (c) presentation speed, (d) self-paced versus experimenter-paced review, and (e) spacing (e.g., the number and scheduling of review presentations). Furthermore, picture review needs to be compared to other forms of review. For example, Koutstaal et al. (1998) found that reviewing a verbal description of events enhanced memory to the same extent as reviewing photographs of the events. Picture review should also be compared to unaided mental review.

As a research tool for autobiographical memory, SenseCam has two important attributes: it can unobtrusively capture a pictorial record of an individual's experiences, and it has sensors that can be set to automatically select the pictures to be taken (as well as providing sensor data). In the present study the current sensor-based trigger algorithm showed little success at selecting the more memorable events. That is, the sensor-triggered picture capture strategy did not have an impact on participants' memory as we measured it when compared to a strategy of capturing pictures at regular intervals and randomly sampling for review and testing. The question of whether a sensor-driven approach to input filtering is appropriate depends on the user's goals and on the extent to which the measured physical environment is systematically related to those goals. Ongoing analyses of the rich data collected here have begun to uncover systematic relationships between sensor values and improved memory performance, and have also begun to demonstrate ways to devise more effective thresholds and selection algorithms (Finley, Brewer, & Benjamin, 2009).

The variability in memory performance that can be accounted for by physical characteristics of the environment almost certainly varies by the structure of peoples' lives, the nature of their activities, their cognitive functioning, and their particular memory goals. Any of these may differ between individuals, and may particularly differ for memory rehabilitation patients versus the healthy young adults in the present study. Thus, adaptive calibration of sensor thresholds and triggering algorithms may be necessary on the individual user level in order to implement effective sensor-based strategies of picture capture.

SenseCam constitutes a new tool for autobiographical memory research that has different capabilities and different disadvantages from those previously available. For example, although SenseCam does provide a veridical visual record which should be useful for evaluating recall accuracy, there are several issues that will have to be resolved: (a) Script-based responses. It is often difficult to differentiate between information in participants' responses that was genuinely recalled versus information that was provided in the picture itself or inferred from the picture and participants' general knowledge of the structure of their own lives (e.g., ''I am walking to Spanish or German class. I know this because the tree angle is very dramatic and it wouldn't be at noon, which is when I'd normally leave this area.''). (b) Narrative summarisation. It seems likely that participants often did not provide as much detail as they actually had in memory (e.g., ''Dinner with friends''). (c) Non-picture information. Many recall responses contain specific details that cannot easily be verified by the picture alone, such as: thoughts or emotions, the topic of a lecture or conversation, the identities of other people, the name of a building (e.g., ''I was in design class being bored out of my mind at this point . . . '', ''telling my boyfriend on aim [AOL Instant Messenger] that all the rum is gone on Friday . . . '', ''I am showing off bad paint brushes that no one washed off so they were useless. I am behind the front desk at [X] Hall . . . '').

Making maximum use of the picture information will require addressing these difficulties. For example the occurrence of script-based responses could be studied by seeing how much the experimenters (acting as control participants) might themselves be able to describe the visual scene or infer what would be pictured next. The issue of narrative summarisation could be addressed by more elaborate instructions or by adjustments in the scoring procedures. The occurrence of non-picture information could be studied using additional experience sampling methods (e.g., Brewer, 1988), or using other information sources (e.g., course lecture notes). Additional research possibilities will continue to arise as additional sensor information becomes available (e.g., audio, location, physiological data).

In this paper we have focused on the use of an end-of-day review of SenseCam pictures to fortify autobiographical memory, and on their effectiveness as memory cues. However, as with any form of external memory, SenseCam pictures could be used for a variety of other purposes (cf. Bush, 1945), such as a tool for reminiscing about the past, for evaluating allocation of activities during a day, to complement human memory by revealing what may not have been noticed or encoded upon initial experience, for back-up reference in case human memory fails, or even to supplant human memory for certain purposes. Proliferation of such technology should continue to transform not just research, but the way we live.

REFERENCES

Bell, G., & Gemmell, J. (2007, March). A digital life. *Scientific American*, *296*(3), 58–65.

Bell, G., & Gemmell, J. (2009). *Total recall: How the e-memory revolution will change everything*. New York, NY: Dutton.

Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., et al. (2007). The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychological Rehabilitation*, *17*(4–5), 582–681.

Brewer, W. F. (1986). What is autobiographical memory? In D. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). Cambridge, UK: Cambridge University Press.

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 21–90). Cambridge, UK: Cambridge University Press.

Brewer, W. F. (1996). What is recollective memory? In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 19–66). Cambridge, UK: Cambridge University Press.

Bush, V. (1945). As we may think. *Atlantic Monthly.*

Conner, T. S., Barrett, L. F., Tugade, M. M., & Tennen, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robins, R. C. Fraley, & R. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pp. 79–98). New York, NY: The Guilford Press.

Conway, M. A. (1990). *Autobiographical memory: An introduction*. Buckingham, UK: University Press.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*, 594–628.

Crovitz, H. F., & Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, *4*(5b), 517–518.

Ebbinghaus, H. (1885). Retention as a function of the number of repetitions. In *Memory: A contribution to experimental psychology* (Ch 6). Translated by Henry A. Ruger & Clara E. Bussenius (1913). New York, NY: Teachers College, Columbia University.

Finley, J. R., Brewer, W. F., & Benjamin, A. S. (2009, October). *Reflections of memory in the environment: Improving the use of physical sensor data to predict memorability of SenseCam pictures from everyday life*. Presentation given at SenseCam 2009 Symposium, Chicago, IL.

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*, 44–66.

Galton, F. (1879a). Psychometric experiments. *Brain*, *2*, 149–162.

Galton, F. (1879b). Psychometric facts. *The Nineteenth Century*, *5*, 425–433.

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Bulter, A., et al. (2006). SenseCam: A retrospective memory aid. In P. Dourish & A. Friday (Eds.), *Ubicomp 2006, LNCS 4206* (pp. 177–193). Berlin, Germany: Springer Verlag.

Jones, W., & Ross B. H. (2007). Personal information management. In F. Durso (Ed.), *Handbook of applied cognition: Second edition* (pp. 471–495). Chichester, UK: John Wiley & Sons, Inc.

Koutstaal, W., Schacter, D. L., Johnson, M. K., Angell, K. E., & Gross, M. S. (1998). Post-event review in older and younger adults: Improving memory accessibility of complex everyday events. *Psychology and Aging*, *13*, 277–296.

Koutstaal, W., Schacter, D. L., Johnson, M. K., & Galluccio, L. (1999). Facilitation and impairment of event memory produced by photograph review. *Memory & Cognition*, *27*, 478–493.

Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376–404). San Francisco, CA: W. H. Freeman.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Schacter, D. L., Koutstaal, W., Johnson, M. K., Gross, M. S., & Angell, K. E. (1997). False recollection induced by photographs: A comparison of older and younger adults. *Psychology and Aging*, *12*, 203–215.

Sellen, A., Fogg, A., Hodges, S., & Wood, K. (2007). Do life-logging technologies support memory for the past? An experimental study using SenseCam. *Conference on Human Factors in Computing Systems, CHI*, Irvine, CA, 81–90.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18*(2), 225–252.

APPENDIX A

Mean and *SD* of participant means of six memory measures by retention interval, review, and trigger condition

	1 week		*3 weeks*		*8 weeks*	
Memory measure	*No review*	*Review*	*No review*	*Review*	*No review*	*Review*
Recollection Judgement						
Sensor	0.41	0.54	0.35	0.38	0.36	0.37
	(0.18)	(0.19)	(0.19)	(0.19)	(0.16)	(0.19)
Timer	0.44	0.52	0.31	0.39	0.28	0.36
	(0.19)	(0.19)	(0.21)	(0.19)	(0.19)	(0.15)
Recognition Rating						
Sensor	3.89	4.14	3.28	3.37	2.97	3.07
	(0.94)	(0.60)	(0.94)	(0.89)	(1.06)	(1.17)
Timer	3.61	4.17	3.11	3.34	2.65	3.22
	(0.86)	(0.78)	(1.02)	(0.95)	(1.03)	(0.92)
Picture-Cued Recall						
Sensor	64.18	68.33	46.25	47.54	38.57	40.44
	(26.52)	(29.21)	(26.33)	(24.27)	(22.03)	(25.88)
Timer	51.56	62.53	38.94	45.42	33.53	39.05
	(24.71)	(24.12)	(23.23)	(25.42)	(22.41)	(26.05)
Time-Cued Recall						
Sensor	78.17	76.38	46.00	55.08	41.29	54.83
	(50.35)	(54.06)	(23.16)	(31.30)	(19.63)	(32.89)
Timer	60.17	84.04	46.96	69.88	55.67	48.00
	(50.02)	(47.40)	(29.25)	(31.79)	(27.08)	(26.32)
Temporal Order Judgement						
Sensor	0.62	0.63	0.63	0.65	0.57	0.64
	(0.09)	(0.09)	(0.08)	(0.08)	(0.11)	(0.09)
Timer	0.63	0.65	0.64	0.64	0.58	0.61
	(0.07)	(0.09)	(0.07)	(0.08)	(0.11)	(0.09)
Recall of Time						
Sensor	1.23	1.27	1.15	1.39	1.11	1.26
	(0.31)	(0.24)	(0.34)	(0.21)	(0.42)	(0.21)
Timer	1.03	1.20	1.10	1.24	1.11	1.08
	(0.26)	(0.29)	(0.30)	(0.32)	(0.39)	(0.33)