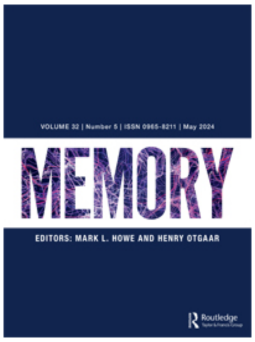

Memory

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pmem20

Accuracy and completeness of autobiographical memory: evidence from a wearable camera study

Jason R. Finley & William F. Brewer

MEMORY
https://doi.org/10.1080/09658211.2024.2377193

Routledge
Taylor & Francis Group

Accuracy and completeness of autobiographical memory: evidence from a wearable camera study

Jason R. Finley [a,b] and William F. Brewer[b]

[a]Department of Psychology, Southern Illinois University Edwardsville, Edwardsville, IL, USA; [b]University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT
A small wearable camera, SenseCam, passively captured pictures from everyday experience that were later used to evaluate the accuracy and completeness of autobiographical memory. Nine undergraduates wore SenseCams that took pictures every 10 s for two days. After one week and one month, participants first recalled their experiences from specific time periods (timeslices), then reviewed the corresponding pictures to make corrections and report information omitted from initial recall. Results demonstrated the utility of wearable cameras as research tools, and illustrated several characteristics of everyday memory. Recall contents reflected the structure of undergraduate lives. Three different types of omissions were reported: neglected, reminded, and forgotten. Pictures stimulated memory, even for non-visual information (e.g., feelings, thoughts), increasing recall by 23%. The mean completeness of initial recall was 79% (upper bound), with at least 21% forgetting. Accuracy was self-scored by participants ($M = 89\%$), and the mean error rate (11%) provided evidence against strong reconstructive and copy theories of memory. The characteristics of errors shed light on the cognitive processes underlying them. Ratings of recall (confidence, reliving, knowledge, and frequency) supported the episodic/semantic distinction, the dual-process theory of repetition, and reconstructive imagery. Metamemory measures showed a positive correlation between confidence and accuracy.

ARTICLE HISTORY
Received 2 August 2023
Accepted 24 June 2024

KEYWORDS
Autobiographical memory; everyday memory; memory errors; SenseCam; wearable camera; pictures

> As fast as the present enters into the past, our states of consciousness disappear and are obliterated. Passed in review at a few days' distance, nothing or little of them remains: most of them have made shipwreck in that great nonentity from which they never more will emerge … –Théodule Ribot, *Les Maladies de la Mémoire, 1881*

How much do we truly remember of our everyday experiences? Such a question is difficult to answer due to the methodological challenges inherent to the study of autobiographical memory. Because of such challenges, the field has lacked much of the basic descriptive data needed to evaluate theoretical claims. The goal of the current study was to develop methodological procedures to provide such data and thus furnish a better understanding of the accuracy and completeness of autobiographical memory. We use the term *autobiographical memory* to cover the broad class of memories relating to the self (e.g., recollective memories, autobiographical facts; see Brewer, 1986). We use the term *recollective memories* for the subclass of autobiographical memories that are typically experienced as the reliving of a specific past event (see Brewer, 1996).

Methodological challenges and strategies in the study of autobiographical memory

To study the accuracy and completeness of any kind of memory, researchers need objective records of the past to compare against participants' responses on a memory test. In most laboratory memory experiments, researchers themselves control and maintain records of the stimuli to which participants are exposed (e.g., word lists). However, in most studies of autobiographical memory (see Woll, 2002 for an incisive review), researchers have little or no control over participants' original experiences, and they typically have limited records of those experiences. For example, a common approach to studying autobiographical memory is the Galton/Crovitz word technique (Crovitz & Schiffman, 1974; Rubin, 2005) in which participants are presented with a word and asked to use that word to generate a recollective memory. While this technique has led to much interesting research (see the review in Rubin, 2002), researchers cannot verify participants' recollections without records of the original experiences, and thus cannot evaluate accuracy and completeness.

CONTACT Jason R. Finley jafinle@siue.edu; jasonrfinley.com Department of Psychology, Southern Illinois University Edwardsville, 1 Hairpin Dr, Edwardsville, IL 62026, USA
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09658211.2024.2377193.

Given the limitations of the word technique, investigators interested in autobiographical memory have developed a variety of alternative procedures. There have been two main approaches to dealing with the problem of verifiability in autobiographical memory research: reliance on records already gathered for other purposes (i.e., unplanned records), and the planned collection of new records for the specific purpose of evaluating memory performance (see Bahrick & Karis, 1982 for an inventory of methods). However they are gathered, such records of experience constitute *external memory*: information stored outside of one's brain (Finley et al., 2018; Finley & Naaz, 2023).

Unplanned records. Whenever some aspects of human experience have been recorded to external memory (e.g., paper, computer, videotape), it is possible for researchers to use such records to query and evaluate participants' memory for those particular aspects of the original experience. Examples of such records that have been used in autobiographical memory studies include: college grades (Bahrick et al., 2008), names of high school classmates and teachers (Whitten & Leonard, 1981; Williams & Hollan, 1981), names of former students (Huang, 1997), sporting event scores (Baddeley & Hitch, 1977), healthcare visits (Jobe et al., 1990; Roberts et al., 1996), news events (Larsen & Plunkett, 1987), names of television shows (Squire, 1989), and the covertly recorded conversations of John Dean, counsel to US President Richard Nixon (Neisser, 1981).

By using such incidentally available information, researchers can evaluate the accuracy and completeness of even decades-old memories, without having had to plan the research many years in advance. However, researchers are then constrained by whatever records happen to be available, which often consist of *autobiographical facts* (e.g., your teacher's name) and lack information on *recollective memories* (e.g., the time a squirrel ripped pages out of your dictionary for nesting material, or when two cockatiels happily squeaked along as you practiced your dissertation defense). Note that some unplanned records do potentially contain more detailed information, for example: incidentally available diaries (Burt et al., 2001; Smith, 1952), letters (Baddeley, 2010), or covert surveillance (Neisser, 1981). Finally, social media in the twenty-first century has provided a new source of unplanned records for cuing autobiographical memories (Talarico, 2022).

Planned records. In order to study memory for personal experiences with somewhat of the detail and systematicity of typical laboratory experiments, it is necessary to collect records for the express purpose of testing and evaluating memory. Several prior studies have done so.

Daily diaries. The first autobiographical memory studies to adopt a planned record approach relied on long-running systematic daily diaries made specifically for studying autobiographical memory (Linton, 1975, 1978, 1982; Wagenaar, 1986; White, 1982, 1989, 2002). In all of these studies, the researchers themselves were the sole participants. The researchers/participants in these heroic studies made structured daily diary records for acquisition periods of 1–6 years, and they used those records to test aspects of their autobiographical memory after retention intervals ranging from one day to 20 years. Later daily diary studies recruited undergraduate and graduate student participants (Barclay & Wellman, 1986; Thompson et al., 1996). For further details on these studies, see Supplemental Materials.

Although such diary methods do enable some objective evaluation of memory for personal experiences (see also Conway, 1990 for a review), they also have several disadvantages. First, there is bias in the selection of which experiences to record in the diary. Second, the very act of making the diary records changes participants' experiences and potentially alters their memories (see Brewer, 1988, p. 82; testing effect aka retrieval practice, Roediger & Butler, 2011; production effect, MacLeod & Bodner, 2017). That said, Thompson's, 1982 "roommate study", which compared recorded events occurring in the life of the participant and events occurring in the life of the participant's (unaware) roommate, suggested minimal memory effects due to writing down the events.

Experience sampling. Advances in technology have allowed the use of experience sampling methods in psychological and medical research (Bolger et al., 2003; Conner et al., 2007; Mehl & Conner, 2012; Stone & Shiffman, 2002). Conner et al. (2007) defined experience sampling as "a collection of procedures that are designed to allow respondents to report their thoughts, feelings, and behaviors over time in natural settings". In an example of experience sampling that inspired the current study, Brewer (1988) had participants carry small electronic "beeper" devices that prompted them at random intervals throughout the day to record certain aspects of their current experience on paper response cards (e.g., date and time, location, thoughts, actions, and various ratings such as location frequency and emotional dimensions). In one experiment, portions of these cards were later used for cued recall (e.g., recalling the *location* of an event when cued with the *thoughts* that were occurring). The beeper methodology offers the advantage of obtaining an unbiased *random sampling* of experiences for later testing. But, like the daily diary method, this method has the disadvantage of frequently and actively involving participants in the capture of information during acquisition, thus potentially altering the very memories to be tested.

Wearable cameras. In the current study, building on prior work by us and by others (Finley et al., 2011; Sellen et al., 2007; Silva et al., 2018; St. Jacques et al., 2011), we made use of a small wearable digital camera (SenseCam; Hodges et al., 2011) to capture pictorial records of the everyday lives of participants. Pictures (i.e., photographs) were taken automatically without any overt signal, so the participants were not aware of which events were being recorded. Thus, unlike previous planned studies of autobiographical memory, our records were not inherently biased in selection, and nor were participants actively engaged in recording their experiences during the acquisition phase of the study.

Accuracy and completeness of memory

Two of the major topics we examine in this study are memory accuracy and memory completeness. Koriat and Goldsmith (1994) referred to these as output-bound versus input-bound measures of memory, respectively. The difference between the two measures is best illustrated by example. Imagine a standard laboratory free recall experiment. Participants study a list of 20 words, then after some retention interval they write down as many of the words as they can remember (free recall). Suppose that a participant writes down 15 words, 10 of which were on the studied list and 5 of which were not. In this case *accuracy* is the proportion of the participant's responses that were indeed on the studied list (i.e., 10/15 = 67%). *Completeness* is the proportion of studied words that the participant in fact later wrote down (i.e., 10/20 = 50%). In the context of everyday autobiographical memory, we can think of accuracy and completeness as answering two basic questions. *Accuracy*: how much of what you remember actually happened? *Completeness*: how much of what actually happened do you remember? Both of these questions have been difficult to answer because of the methodological challenges reviewed above. This in turn has impeded resolution of theoretical controversies.

Theories of memory accuracy

Views on the accuracy of recollective memories have ranged from *copy theories* at one extreme to *strong reconstructive theories* at the other. Copy theories view recollective memories as accurate copies of past experiences. For example, Penfield wrote that stimulating the cortex with an electrode appeared to replay original experience: "The record apparently includes all that the individual was aware of at the time" (1969, p. 165). Brown and Kulik (1977) theorised that a triggering "flashbulb" event could activate a mechanism to "record permanently all immediately previous and contemporaneous brain events" (p 87). They stated that flashbulb memories are "very like a photograph that indiscriminately preserves the scene" (p. 74).

In contrast, a number of theorists have instead argued that recollective and flashbulb memories are strongly reconstructed in ways that cause them to be very inaccurate representations of the original experience (e.g., Barclay, 1986; Loftus & Loftus, 1980; Neisser, 1982, pp. 43–48; Neisser & Harsch, 1992). Barclay (1986) reviewed the limited evidence on the accuracy of everyday autobiographical memory and concluded that "when information is remembered, acquired autobiographical self-knowledge drives the reconstruction of plausible, but often inaccurate, elaborations of previous experiences. Memories for most everyday life events are, therefore, transformed, distorted, or forgotten" (p. 89).

Between the two extremes, Brewer (1986) proposed a *partially reconstructive theory* that recollective memories are "reasonably accurate copies of the individual's original phenomenal experience" (p. 43) but that they are subject to a variety of reconstructive processes that can cause errors, just as are other forms of memory (see reviews in Roediger & DeSoto, 2015; Schacter, 1995).

Although copy theories in particular are no longer widely endorsed, it is still worth considering the entire range of theories on memory accuracy because there has been little to directly test them in the context of everyday experiences.

Episodic versus semantic memory

Tulving (1972) first proposed a general distinction between memory for specific episodes (*episodic*) versus knowledge of the world (*semantic*). Brewer (1986) later distinguished between *episodic autobiographical memories* ("personal memories", "recollective memories") which are typically experienced as a reliving of the original episode with "mental time travel" (Suddendorf & Corballis, 2007; Tulving, 2005), and *semantic autobiographical memories* which consist of general knowledge about the self. Brewer's distinction gained some force by analogy with the episodic/semantic distinction as used for laboratory tasks, yet lacked the support of empirical evidence from the domain of autobiographical memory. Later evidence was provided by Levine et al. (2002) who developed an Autobiographical Interview to distinguish episodic (internal details) from semantic (external details) components of autobiographical recall. In the current study, we offer an additional technique for teasing apart episodic and semantic memory in everyday life, and for characterising their differences and cooccurrences.

Current study

Our overall goal with this study was to provide better descriptive data about the basic characteristics of everyday autobiographical memory, including estimates of accuracy and completeness, that can constrain theoretical claims such as those just reviewed. The use of wearable cameras in the current study enabled us to avoid the problem of non-random sampling of experience, and the problem of participants being actively involved in making records. In addition to the use of the camera technology, our method included gathering a variety of rating data during the later memory tests (e.g., confidence, reliving vs. knowledge) that were intended to speak to a number of other theoretical issues such as the distinction between episodic and semantic memory, and metamemory for everyday experiences.

Method

Participants

Participants were nine undergraduates at the University of Illinois at Urbana-Champaign, three men and six women, ages 18–22 ($M = 20.44$, $SD = 1.17$). They participated in exchange for $35 paid in instalments of $10 at the end of the acquisition period, $10 at the first testing session, and $15 at the second testing session. The entire study took place during the summer of 2010. Participants were healthy and had no memory impairments. Although this sample size may seem small by comparison to basic laboratory studies, it is nine times larger than the single-participant studies that used the daily diary method (Linton, 1982; Wagenaar, 1986; White, 2002), and equivalent to the sample sizes used in Brewer's, 1988 beeper study ($N = 8$ in Experiment 1, and $N = 10$ in Experiment 2).

Recruitment and selection. Participants were recruited via posters and internet notices. Roughly 100 individuals expressed initial interest via email and were sent a full description of the procedure and a link to an online questionnaire. A total of 23 individuals completed this questionnaire, which covered basic demographic information, living situation (e.g., number of roommates), experience with technology, and anticipated daily activities for the summer (e.g., classes, work, sports). Finally, nine participants were selected on the basis of their availability, apparent reliability, and having living situations and daily activities that would least interfere with wearing a SenseCam for a considerable portion of two non-weekend days.

Privacy. Several aspects of the 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 email message they received. Throughout the study, each participant's data were identified only by a unique code number assigned to them. Pictures from an individual participant's acquisition period were never seen by anyone except the researchers and that participant. Pictures were stored on an external hard drive kept in a locked cabinet in the lab.

Participants could deactivate the SenseCam (using the on/off button or the privacy button) at any time to avoid the possibility of pictures being captured. Pressing the privacy button deactivated picture capture for 7 min, with a warning beep 15 s 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. In addition, when entering their workplace or a private residence, participants were instructed to deactivate the camera until they received permission to reactivate it.

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 the acquisition period without having been seen by anyone. Four of the nine participants made use of this technique, and their durations for deletion were: 5 min, 7 min, 1 hr, and two 20 min periods.

Finally, during the testing sessions, participants were instructed that if they remembered anything that they considered private, they could just type "private" instead of providing a recall. An overview of the ethical issues involved in using a wearable camera for research is provided by Kelly et al. (2013). The current study received ethical approval from the Institutional Review Board of the University of Illinois at Urbana-Champaign, protocol number 08229. Participant instructions, informed consent form, and debriefing can be found in Supplemental Materials.

Apparatus

SenseCam is a small, wearable, sensor-equipped digital camera initially developed by Microsoft Research Cambridge (Hodges et al., 2011). It does not capture video or audio recordings; it only captures still pictures. Commercially available descendants of the original SenseCam device have included Vicon Revue and Autographer. The SenseCams used in this study were hardware version 2.3b and firmware version 2.6.7. Pictures were captured using a fisheye lens with a wide angle of view (119° diagonal) meant to approximate the human field of view better than traditional lenses. A built-in adjustable lanyard allowed SenseCam to be worn around the neck at upper chest height, resulting in pictures with a field of view similar to that of the participant's actual perceptual experience. The pictures were stored with timestamps on a 1Gb SD flash memory card and saved as JPEG files with display size 640×480 pixels, resolution of 72 dots per inch, colour depth 32-bit (16.7 million colours), and file size of approximately 30 kilobytes each. The maximum rate of picture capture was approximately once every 5 s. The SenseCam can be set to capture pictures in response to changes in sensor values, and/or at fixed intervals based on a timer. We used the timer setting in this study, as it would be more reliable and because in a previous study, we found no difference in memory effects of pictures that were sensor-triggered versus timer-triggered (Finley et al., 2011). We wanted to capture as many pictures as possible, but pilot testing found that using the maximum capture rate (once every 5 s) drained the battery too quickly. Thus we set the capture rate to once every 10 s, to strike a balance between capture frequency and battery life. Additional information about SenseCam can be found in Finley et al. (2011) and detailed technical specifications can be found in Hodges et al. (2006).

The device and others like it have been used in research aimed at assisting people with memory impairments as well as other health-related research (e.g., Doherty et al., 2013; Laursen, 2009; Silva et al., 2018; Woodberry et al., 2015). People with unimpaired memory who have used SenseCam have anecdotally reported that watching a series of one's own SenseCam pictures can invoke a vivid sense of reliving, including stimulation of memory for non-visual aspects of experience (Berry et al., 2006; Loveday & Conway, 2011b). Wearable camera images have also been used in conjunction with brain imaging techniques such as fMRI to explore the neural mechanisms of autobiographical memory retrieval (Chow & Rissman, 2017). Social and personality psychology uses have also emerged (Brown et al., 2017). Commercially, wearable cameras have gained niche prominence as body-worn cameras used by police departments (Ariel et al., 2015), and as used for outdoor adventures and sports (e.g., GoPro; Vannini & Stewart, 2017). The automatic nature of wearable cameras avoids any potential memory effects of manually taking pictures (Henkel, 2014).

Participants also carried a small paper notepad that they could use to note any time periods for deletion if they forgot to deactivate the camera. The notepad also contained a prepared statement about the study for anyone who had questions, reminders about how to handle private locations, reminders on how to operate the SenseCam, and the researcher's contact information.

Design

Participants wore a SenseCam for an acquisition period of two consecutive weekdays. Participants then returned to the lab for two test sessions: roughly 1 week and 1 month after acquisition. In a given test session, participants were tested on their memory for only one of their two acquisition days (order counterbalanced across participants), to avoid rehearsal of the events from the other acquisition day. Each acquisition day had been segmented by the researchers into timeslices (described in detail below), and the timeslices for a given day were tested in a random order.[1]

We chose to sample weekdays because they are more representative of everyday life, both in proportion (5/7) and in routine.

Procedure

Our methodological goals in this study were to develop procedures that: (a) used a wearable camera to *unobtrusively* gather pictures of an *unbiased* sample of experiences from the everyday lives of our participants; (b) allowed us to cue and test memory for these experiences without giving away too much information to the participants; (c) encouraged the participants to recall and rate as much as possible about their experiences; (d) allowed the participants to use the pictures to score the accuracy of their recalls; and (e) avoided the aforementioned ethical issues surrounding privacy. To meet these goals it was necessary to develop a somewhat complex procedure, which we will now describe in detail.

Acquisition period. On the day before the start of the acquisition period, participants were trained in the use of the SenseCam and completed a form detailing their anticipated schedule for the upcoming days. Participants were also informed that their memory for the events occurring during the acquisition period would be tested later, 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.").

During the acquisition period, participants wore a SenseCam as they went about their normal daily activities for two consecutive weekdays, recharging the SenseCam each night. The starting and ending time of acquisition for each day were determined by each individual participant, and thus varied by participant and by day. The participants typically began wearing the camera after they were dressed and ready to start their daily activities; they typically stopped wearing it as they were about to go to bed. Across all participants and days, the mean starting time was 11:00 AM ($SD = 2$ hr 33 min), and the mean ending time was 10:19 PM ($SD = 2$ hr 19 min). The actual mean amount of time that the camera was active per day (adjusting for periods of deactivation or deletion) was 9 hr 7 min ($SD = 3$ hr 34 min, range: 1 hr 28 min to 15 hr 59 min). The mean amount of time per day that the camera was disabled or pictures were deleted was 2 hr 11 min ($SD = 2$ hr 14 min, range: 0–7 hr 38 min). The rate of picture capture was set to once every 10 s. Across all participants and days, the empirically derived mean time interval between picture captures was 10.4 s ($SD = 1.3$). The mean number of pictures taken per day was 3,171 ($SD = 1{,}274$, range: 437–5,497). Participants returned the SenseCam and notepad to the lab within several days of the end of the acquisition period. The researchers deleted any pictures falling in a time period marked for deletion in the notepad, and copied the pictures from the camera onto an external hard drive.

Timeslice segmentation. In order to test participants' recall, we needed to cue them to a particular set of events. In our first SenseCam study (Finley et al., 2011), at the end of the procedure we pilot tested open-ended free recall of an entire day. Results showed this approach to be too unconstrained, leaving too much room for variance in the amount of time participants summarised over. Thus for the current study, we segmented each day into conceptually coherent and verbally describable *timeslices* for testing. We did this by examining all of a participant's pictures captured on a given day, supplemented by the participant's anticipated schedules collected before acquisition, in search of *episode transitions* which could be verbally described by the researchers without giving away too much information during testing.[2] Our goal in creating timeslices was to describe them such that a hypothetical third-party observer of the pictures would be able to identify the subset series of pictures corresponding to the verbal description. For example, one verbal description of a timeslice, along with date and time information, was:

> 1st day you wore SenseCam. Thursday July 1, 2010. 2:29:03 PM to 2:43:47 PM. FROM the moment that you walked out of the building where you had just had a meeting, UP UNTIL the moment that you later sat down outside with a group of people.

The mean number of timeslices our procedure produced per participant per day was 11.4 ($SD = 3.4$, range: 5-18). Not all timeslices were ultimately tested, due to time limits in the memory test sessions. For those timeslices that were tested, the overall mean of the participant mean of the median timeslice duration per day was 37.4 min ($SD = 8.4$). The minimum tested timeslice duration was 3.8 min and the maximum was 300.5 min (i.e., 5 hr). Additional basic descriptive data about timeslices is available in the Supplemental Materials, Table S1.

Memory tests. Participants returned to the lab for two testing sessions, at retention intervals of approximately 1 week and 1 month. We chose these intervals in part because a week and a month are naturally occurring units of time that are common in everyday life. But also we chose these intervals to guard against possible floor or ceiling effects, which could have dramatically reduced our usable data. We wanted to cast a net so as to catch a level of performance somewhere between those two extremes. The mean number of days that the testing sessions occurred after the final acquisition day was 7.7 days ($SD = 0.9$, range: 6–9) for Test 1, and 28.9 days ($SD = 3.5$, range: 22–36) for Test 2.

Tests were administered individually on computers, with all instructions presented via onscreen text. The entire testing interface was programmed by the first author specifically for this study, using REAL Studio 2010 (formerly known as REALbasic, subsequently known as Xojo). Programming a custom interface (Figures 1 and 2) was essential in allowing us to gather the rich data we sought.

Each test covered one of a participant's two acquisition days, with test order counterbalanced across participants. Participants were tested on a particular acquisition day's timeslices in a random order. For each timeslice, there was a three-phase procedure: (a) initial recall; (b) picture review; and (c) self-scoring. All three phases were completed for a given timeslice before the test advanced to another timeslice.

Initial recall (phase 1 of 3, per timeslice)

Timeslice cues. Recall of each timeslice was cued with time information and a brief researcher-developed verbal description that attempted to capture the beginning and end point of the period of time. Figure 1 shows an example of the initial recall testing screen. No pictures were shown during initial recall.

Response category cues. Participants were instructed that their task was to describe everything about their original experience during the timeslice in as much detail as they could remember, and that they should describe things even if they were not completely certain about them. In order to emphasise our desire that the participants provide us with everything that they could remember about a timeslice, we provided exhaustive category cues for each of the major aspects of experience that are typically found in an autobiographical event (see Barsalou, 1988; Brewer, 1986, p. 34, 1988, pp. 74–75; Johnson et al., 1988; Rubin et al., 2003). Our choice of category cues was also influenced by the range of responses participants gave in a whole-day free recall pilot phase at the end of our earlier SenseCam experiment (Finley et al., 2011). Below the verbal description of the timeslice being tested, a list of response category names and descriptions was displayed in a fixed order. Table 1 shows the list of categories and their descriptions. The overall instructions to the participants read:

> You will type everything you can remember into response rows organized into a list of the following categories: … For each category, please type every distinct element that you remember into a different response row. … You will be able to add as many response rows as needed for each category.

Participants' mean output order (response number per timeslice) ended up being almost perfectly correlated with the order of the categories in the interface, $r = .99$, $p < .001$.

Entering and rating category elements (response rows). The interface allowed the participants to add one or more response rows for each category and the participants were instructed that separate response rows should be used for every distinct category element that they remembered for a given category for the timeslice being tested (e.g., two response rows if they remembered two different locations). This procedure led the participants to break down the longer and more complex timeslices into smaller more discrete elements so that we could focus our various rating scales on the simpler, less complex elements. There was no time limit for entering response rows. On average, participants entered 12 response rows per timeslice ($SD = 6$, range: 3–22), or 30 responses per hour of original experience ($SD = 17$, range: 6–65).[3]

Each response row consisted of a main text box in which participants typed a response (e.g., for *location*: "The corner gas station"). Next to the main text box there were: four drop-down menus for rating the response on *confidence*, *reliving*, *knowledge*, and *frequency*; and an additional text box in which participants could optionally type any notes clarifying their ratings.

The *confidence rating* asked participants how certain they were that this aspect of original experience actually

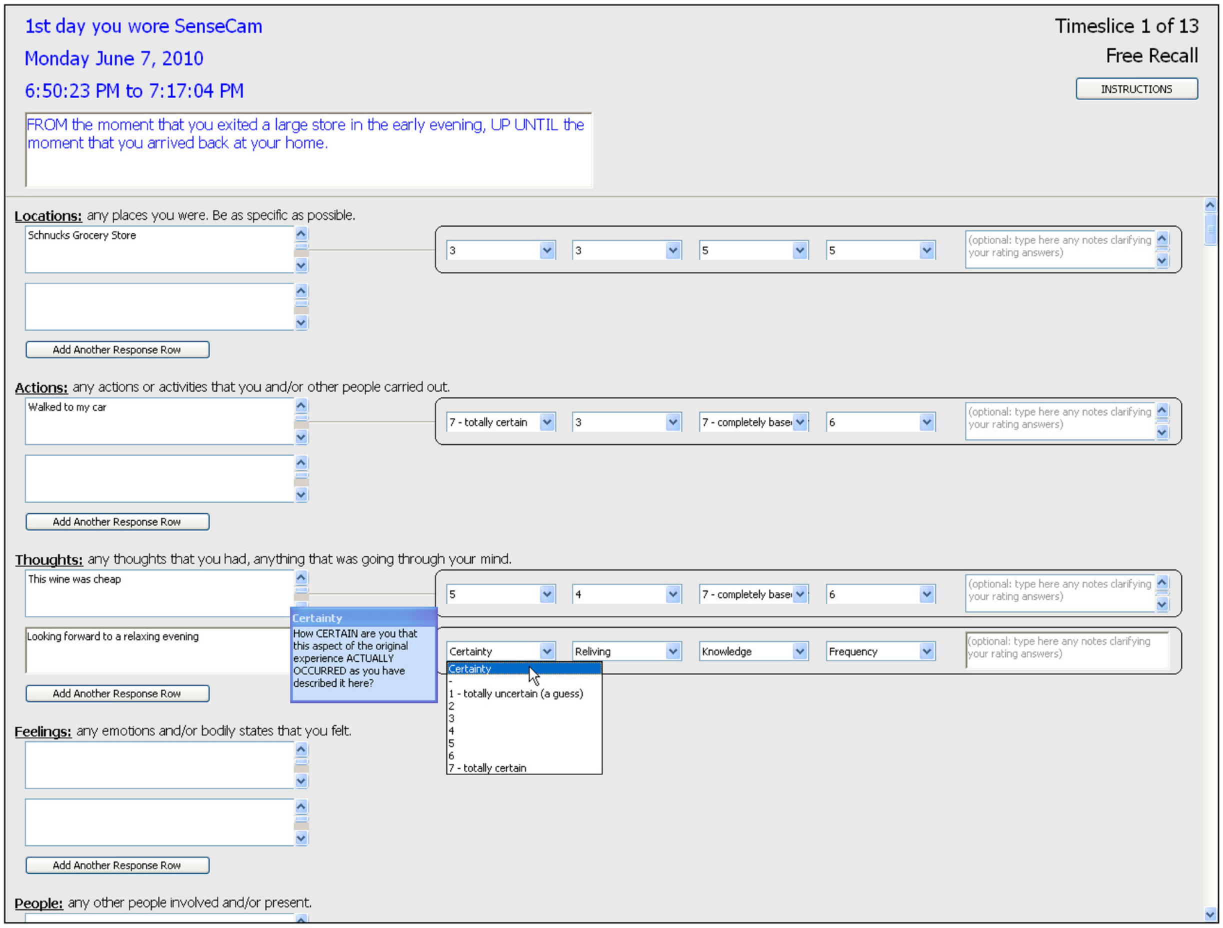

Figure 1. Screenshot of initial recall phase for a timeslice.

Note: Participants could scroll down to see all of the categories (as listed in Table 1). Two response rows were provided by default for each category, but participants could click a button to add more rows as needed.

occurred as they had described it in the row's main text box. The rating was made on a 1–7 scale, where 1 was labelled as *totally uncertain (a guess)* and 7 was labelled as *totally certain*. This scale was a retrospective confidence judgment, and for the participants it was actually labelled "certainty" but we will refer to it as "confidence" throughout this paper.

The *reliving rating* asked participants the extent to which they felt as if they were reliving this aspect of the original experience as they remembered it. The rating was made on a 1–7 scale, where 1 was labelled *not at all* and 7 was labelled *as clearly as if it were happening right now*. This scale was intended to measure the quality of the participants' recollective experiences (cf. Brewer, 1988, pp. 44–45; Rubin et al., 2003).

The *knowledge rating* asked participants to what extent their description of this aspect of the original experience was based on their background knowledge rather than recollection, and included the following clarifying text: "Sometimes people know or think that something happened without being able to actually remember it. They may know or infer that something happened based on their knowledge about the world, their lives, or their personal routines/schedules". The rating was made on a 1–7 scale, where 1 was labelled *not at all based on knowledge* and 7 was labelled *completely based on knowledge*. This scale was intended to measure the participants' beliefs about the degree to which a given answer was based on their general knowledge about their everyday lives (i.e., semantic autobiographical memory) instead of being based on a specific episodic recollection (i.e., episodic autobiographical memory).

The *frequency rating* asked participants to judge how frequently this aspect of the original experience occurred in their everyday lives. The rating was made on a 1–7 scale, where 1 was labelled as *very infrequently* and 7 was labelled as *very frequently*. This scale was intended to measure event/location frequency which has been shown to play a major role in the forgetting of autobiographical memories (Brewer, 1988; Linton, 1982; Wagenaar, 1986; White, 1982).

Overall timeslice ratings. Once participants finished entering and rating response rows for the timeslice being tested, on a new screen they made 5 *overall* ratings for the *entire timeslice* on 1–7 scales: memory strength, sense of reliving, perspective of remembering

Table 1. Response category descriptions.

Response category	Description provided in instructions
Locations	Any places you were. Be as specific as possible.
Actions	Any actions or activities that you and/or other people carried out.
Thoughts	Any thoughts that you had, anything that was going through your mind.
Feelings	Any emotions and/or bodily states that you felt.
People	Any other people involved and/or present.
Information Content	Any information (e.g., topics, facts, stories, etc.) contained in any conversations, readings, writings, lectures, computer usage, video, radio, etc. Be specific.
Visual & Spatial	Descriptions of anything that you saw and any aspects of the visual scene, and where things were spatially situated in relation to you or to each other.
Audio	Anything that you heard (e.g., sounds, voices, music, noise, etc.).
Touch, Taste, & Smell	Anything you felt, tasted, or smelled.
Clothes & Belongings	Clothes you were wearing, and any belongings that you had with you and/or were using.
Environment/ Weather	Anything such as temperature, brightness, humidity, air movement, precipitation and sky clearness/ cloudiness (if outside), etc.
Etc.	Any aspect of anything you remember that does not fit in the other categories.

Note: Categories are listed here in the same order that they were provided in the initial recall and self-scoring phases of the test. Response counts and mean ratings for each category are given in the Supplemental Materials, Table S3.

(1st- vs. 3rd-person), frequency of having thought or talked about the original experience, and how much more they thought they would remember upon viewing the timeslice's pictures. This attempt to gain information about the entire timeslice ended up adding very little to the more specific and detailed category responses and so will largely go unreported in this paper, only appearing in Memory Increase Due to Pictures and Supplemental Materials Table S1.

Picture review (phase 2 of 3, per timeslice)

Picture review. After the initial recall phase for a given timeslice, participants were instructed that they would now see a movie of the timeslice's pictures, and were reminded that they would only be shown pictures that were taken by the SenseCam while they were wearing it (i.e., they would not be shown anyone else's pictures). All of the pictures from the timeslice being tested were presented in chronological order at the rate of 6 per s. Given that pictures were captured approximately every 10 s, that meant that 1 hr of original experience corresponded to about 1 min of picture review. The experience of viewing such a sequence of images is somewhat like watching a jerky fast-forwarded replay of a video shot from the first-person perspective (relatedly, see rapid serial visual presentation, Spence & Witkowski, 2013). Participants could not pause, fast-forward, or rewind the picture presentation during this phase. A bar beneath the pictures showed progress through the sequence of pictures in the timeslice. The durations of the review ranged from 4 to 301 s. Participants were instructed that they would be able to review the pictures at their leisure in the next phase when they self-scored their initial responses for the timeslice being tested. The rapid presentation rate was used to ensure that every participant viewed all images, without taking up too much of the testing session.

Success of timeslice descriptions. The pictures shown in the picture review were always indeed those that correctly corresponded to the *timeslice cues* given in the initial recall phase (date, times and verbal description). However, we foresaw the possibility that during initial recall the timeslice cues alone might have failed to get participants thinking about the particular segment of original experience that we wanted them to remember. For example, they might have been thinking of a different episode that occurred earlier or later in that day. Thus, immediately after the picture review, participants answered a yes/no question as to whether the segment of original experience that they were thinking about during initial recall was indeed the same as that shown in the pictures. This question was included to check the success of the researcher-developed verbal timeslice descriptions in cueing the participants to the specific timeslice we had intended. To preview, the mean success rate was 87%.

Self-scoring (phase 3 of 3, per timeslice)

One of the goals of this study was to investigate the accuracy of autobiographical memory for everyday experiences in the lives of our participants. In a standard laboratory experiment on human memory the researcher knows much more than the participants about the to-be-remembered information. However, in a study of autobiographical memory, the reverse is true. For example, suppose that during initial recall a participant stated: "Bonnie ate lunch with me. She was sitting across the table". If the corresponding pictures show no one sitting across the table during lunch, then both the researcher and the participant would be able to score the recall as an error. However, if the pictures do show a person sitting across the table, the researcher has no way to tell if that person is in fact Bonnie or perhaps someone else (e.g., Erica). Only the participant themself has the information needed to score their initial recall as accurate versus error. Thus we developed a procedure to allow the participants to score their own recalls, and we used these scores for all of our analyses involving memory accuracy.

Participants were instructed that they would be making corrections to the responses they made in the initial recall phase for this timeslice, and adding new responses (i.e., reporting information omitted from initial recall). Figure 2 shows an example of the *self-scoring* screen. The acquisition day, date, and time of the timeslice being tested were again displayed at the top of the screen, along with the researcher-developed verbal description of the timeslice. In addition, a movie player-like interface allowed participants to further review any pictures from the timeslice being tested. In the lower portion of the screen, the list of response categories was again displayed, along with the corresponding text from all of the participant's initial recall responses for the timeslice being tested, and space to add new responses.

Self-scored accuracy of initial recall responses. Now having once viewed all the pictures from the timeslice being tested, and having the ability to further review all of those pictures at their leisure, the participants rated the accuracy of each of their initial recall responses (i.e., a response row for a given category cue) from 1 to 7 (1 = *totally inaccurate*, 7 = *totally accurate*) and rated their confidence in their accuracy rating from 1 to 7 (1 = *totally uncertain about the accuracy rating (a guess)*, 7 = *totally certain about the accuracy rating*). Whenever an initial response was rated as less than totally accurate (i.e., any rating less than 7 out of 7), participants were required to enter text correcting their initial response. Worth noting, participants showed themselves willing to self-score with the full accuracy scale, with three of them using all 7 scores at least once, five using 6 scores, and one using 5 scores. This demonstrates the viability of such self-scoring.

Omissions (reported after pictures). Participants could add one or more new response rows for each category cue (see Figure 2). Each new response row included the same components as the response rows in initial recall: a main text box; ratings on confidence, reliving, knowledge, and frequency; and an optional text box for clarifications. We will refer to these added responses as *omissions (reported after pictures)* or simply *omissions*, because they represent information about the original experience that was omitted from initial recall, and because omission is the term used in traditional laboratory list learning experiments to indicate an item the participant studied but did not output during a recall test. We did not use this terminology with participants; the instructions to participants in the interface simply referred to adding new responses.

We believed that the overall class of *omissions* could include responses that derived from different psychological processes. Therefore we asked our participants to use a new drop-down menu to classify each such new response into one of the following four types: (Neglected) *I DID remember this in Free Recall but didn't write down*, (Reminded) *I DID NOT remember this in Free Recall and now I DO remember it*, (Forgotten) *I DID NOT remember this in Free Recall and I STILL DO NOT remember it*, or (Other) *Other (please explain in text box to the right)*. The Forgotten option was included to accommodate responses about aspects of original experience that were

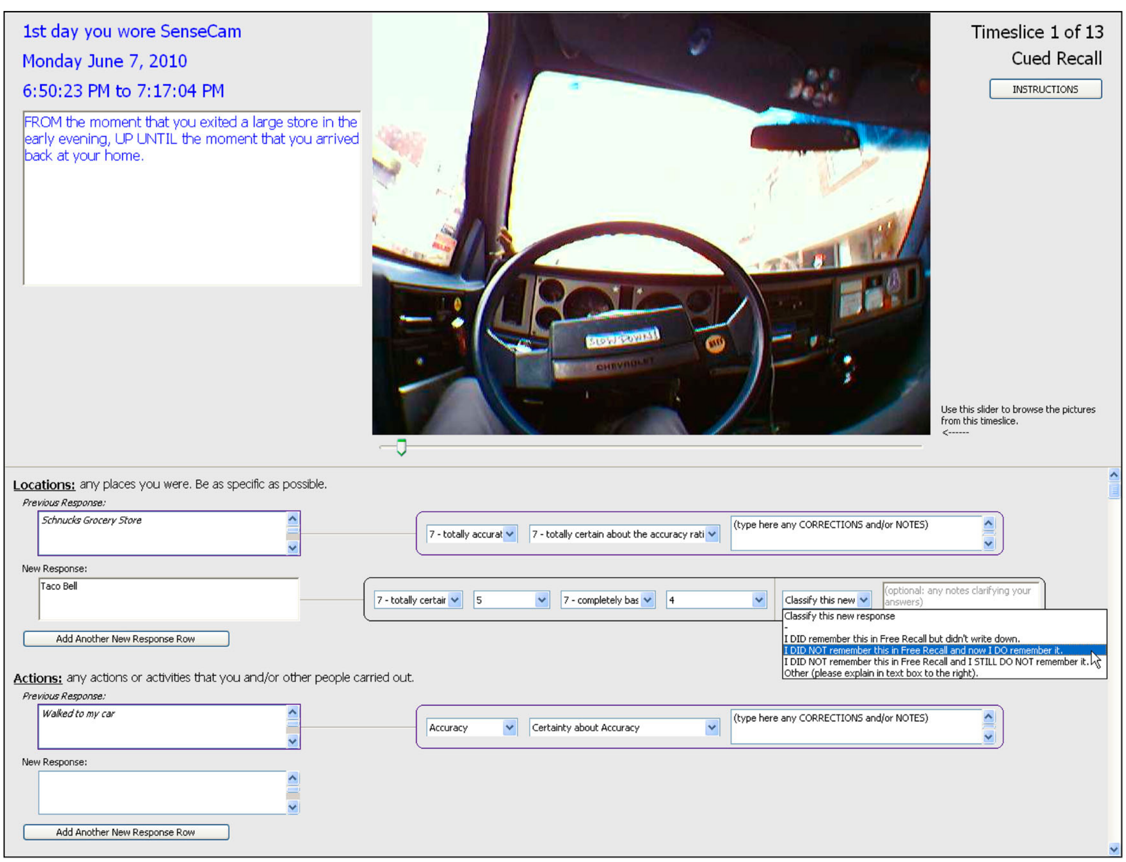

Figure 2. Screenshot of self-scoring phase for a timeslice.

Note: Assisted by the pictures from the timeslice, participants scored the accuracy of their initial recall responses (example: Schnucks Grocery Store, see Figure 1) and reported anything omitted from initial recall (example: Taco Bell). The picture used as an example in this figure comes from the first author's use of a SenseCam.

clearly evident in the pictures but that participants still did not remember.

Overall timeslice ratings. Finally, once participants finished self-scoring and reporting omissions, they again made five overall ratings for the timeslice. These ratings questions were essentially the same as those that followed initial recall, and similarly added little to the more specific and detailed category responses and so will only appear in Memory Increase Due to Pictures and Supplemental Materials Table S1.

Subsequent timeslices. Participants completed all three phases (initial recall, picture review, and self-scoring) for a given timeslice before moving on to another timeslice. Participants completed the three phases for a series of timeslices from one acquisition day, in a random order, until they completed all of that day's timeslices or reached the 2 hr time limit. They were allowed to take breaks between timeslices. On average participants completed 76% of the day's timeslices within the time limit ($SD = 21\%$, range: 33–100%). Participants were debriefed at the end of the second testing session.

Results and discussion

The Results and Discussion section is organised as follows: (a) steps taken to prepare the data for analysis, (b) the contents of autobiographical memories, (c) types of omissions reported after viewing pictures (neglected, reminded, forgotten), (d) nonvisual remindings, (e) completeness of autobiographical memory, (f) comparable retention curves in autobiographical memory, (g) memory increase due to pictures, (h) accuracy of autobiographical memory with a focus on types of errors, (i) episodic versus semantic autobiographical memory, and (j) metamemory. We ran this study with the aim of broad relevance to a number of issues in autobiographical memory. Thus, we will discuss theoretical implications and related prior studies alongside our results to put them in context.

An alpha level of .05 was used for all tests of statistical significance. Standard deviations are reported raw (i.e., calculated using N, not N-1). Effect sizes for comparisons of means are reported as Cohen's d calculated using the pooled standard deviation of the sets of values being compared. Correlations, except where otherwise noted, were calculated as repeated measures correlations (r_{rm}), which account for individual differences without sacrificing power (Bakdash & Marusich, 2017). These correlations are reported with degrees of freedom in parentheses or in table notes. In example quotes from participant responses, all text is preserved exactly as entered by participants (including any mistakes in spelling, grammar, punctuation, or capitalisation), with the exception that any names of people are replaced with bracketed initials to preserve privacy.

Data preparation

Miscued timeslices. Participants responded to a total of 150 timeslices. After the picture review for a given timeslice, participants were asked if the segment of their lives that they had just attempted to recall was indeed the same as that shown in the pictures. They said no to 25 timeslices, indicating that for those timeslices our verbal description along with time and date did not successfully cue participants to the intended segment of their original experience. We excluded the 25 miscued timeslices from all analyses. Averaged across participants, such miscued timeslices comprised only 13% of tested timeslices ($SD = 20\%$, range: 0–56%). In other words, our technique for generating verbal cues for the timeslices was successful on average 87% of the time ($SD = 20\%$, range: 44–100%). The success rate was 100% for five of our nine participants. Future studies could likely achieve even greater success by testing timeslices in chronological order rather than random order.

Retention interval. Our primary reason for including two retention intervals in this study (approximately 1 week and 1 month) was to safeguard against possible floor or ceiling effects in the initial recall data. Examination of the data showed no obvious floor or ceiling effects so we collapsed the data across the two retention intervals in all subsequent analyses. That is, for each participant we combined data points from both testing sessions and proceeded with analysis as if there had been only one testing session.

Although our measure of memory completeness declined slightly from test 1 to test 2, the difference was not statistically significant. Nor did retention interval have a statistically significant effect on our measure of accuracy or any of the initial recall ratings (confidence, reliving, knowledge, and frequency). Thus we are not losing meaningful information by collapsing across retention intervals. We further discuss the issue of retention interval, and report our results separately for test 1 and 2, in Supplemental Materials, including Tables S1 and S2.

Potentially visible response categories. The response categories differed in how much information the pictures provided for scoring recall accuracy and for cueing omission reports (to be used for analysing completeness). For example, the pictures almost always provided evidence about *location* but rarely provided direct evidence about *thought*. Thus, we split the data into response categories for which the pictures were most likely to provide evidence for self-scoring of memory accuracy (Actions; Clothes & Belongings; Environment/Weather; Information Content; Locations; People; and Visual & Spatial) and those categories for which the pictures were less likely to provide such evidence (Audio; Feelings; Thoughts; and Touch, Taste & Smell). For brevity, we will refer to the former set of categories as *visible* and the latter as *nonvisible*. Initially one might think *information content* should be classified as nonvisible, however in practice the pictures often showed computer screens, open books, or classroom visual displays that could be used to score the recalls of information content.

Because the scoring of accuracy and the reporting of omissions were contingent on the visual information about the original experiences being available in the pictures, subsequent analyses will be based only on data from the potentially visible response categories unless otherwise noted. Across all participants, there was a total of 1,035 responses from the visible categories and 375 responses from the nonvisible categories. Our decision to distinguish between visible and nonvisible categories was supported by the fact that participants' confidence about the accuracy ratings they gave to their initial recalls was statistically significantly higher for the visible categories ($M = 6.10$, $SD = 0.58$) versus the nonvisible categories ($M = 5.73$, $SD = 0.96$), $t(8) = 2.51$, $p = .036$, $d = 0.44$.

Contents of autobiographical memories

One of the very first accounts of the contents of autobiographical memories was Brown and Kulik's (1977) paper on flashbulb memories. Brown and Kulik found that flashbulb memories typically contained information about: *place*, *ongoing event*, *own affect*, *informant*, *affect in others*, and *aftermath* (see Brewer, 1992, pp. 278–280 for a detailed analysis of these findings).

Brewer (1988, p. 77) argued that the contents of recollective memories are driven by the content and structure of the underlying *event* originally experienced by the individual recalling the event. He concluded that, given the structure of the everyday lives of American college students, a typical undergraduate recollective memory should contain observable information about *locations*, *actions*, and *people*, and unobservable information about *thoughts* and *feelings*.

Additional data relevant to these claims can be found in Figure 3. This Figure shows the mean proportion of participants' initial recall responses belonging to each response category, collapsed across testing sessions then averaged across participants, with the potentially visible categories shown as grey bars.[4] The frequency of responses for the visible categories is about twice that for the nonvisible categories. The most frequent visible categories were *locations*, *actions*, *clothes & belongings*, and *people*. The most frequent nonvisible categories were *feelings* and *thoughts*. The recall of these internal aspects of original experience contradict Ribot's claim that nothing or little of them would remain after several days (1881, p. 46., via James, 1890, p. 680). Our data are in nice agreement with Brewer's earlier data and his account of the content of recollective memories based simply on the ecology of the lives of typical American undergraduates. Mean ratings and response lengths as a function of response category (both visible and non-visible) are available in the Supplemental Materials, Table S3.

Table 2 shows the overall correlations among the response ratings in initial recall, which will be relevant for a number of subsequent analyses.

Three different types of omissions (Reported after pictures)

During the self-scoring phase, once participants had seen their pictures and still had the pictures on the screen, they could report information that had been omitted in initial recall, and classify these omissions as: neglected, reminded, or forgotten. For example, suppose a participant is now looking at a picture showing a coffee cup in their hand, but they had omitted any mention of the cup during initial recall. They now report that they were holding a coffee cup. If they classify that report as *neglected*, that means that they recalled the cup before seeing the picture, but just neglected to mention it during initial recall. If they classify the report as *reminded*, that means that they did not recall the cup before seeing the picture, but now they do. If they classify the report as *forgotten,* that means that they did not recall the cup before seeing the picture, and they still have no memory of it.

Table 3 shows the ratings and counts for the three different types of omissions (*neglected*, *reminded*, and *forgotten*).[5] The data in Table 3 are for the potentially visible response categories only (i.e., excluding feelings, thoughts, audio, and touch/taste/smell). Table 3 also includes, for comparison, the ratings for initial responses that participants self-scored as accurate (7 on the scale of 1–7). In the Supplemental Materials, we analyse the rating profiles of the omissions, and consider them in light of previous research, providing some insights into the different memory processes associated with each type of omission. We think it is important to be able to distinguish between these three qualitatively different types of omission reports evoked by the pictures, and these distinctions will also prove useful in our later calculation of a memory completeness measure. Any future studies that elicit recall both before and after viewing photos or videos should include such a distinction in their procedure.

Nonvisible remindings

To what extent did viewing the SenseCam pictures remind participants of even nonvisual aspects of the original experience? Although we chose to exclude from most analyses responses from nonvisible response categories (feelings, thoughts, audio, touch/taste/smell) on the grounds that their accuracy could not be determined from pictures, we return now to briefly consider these categories in this section. Of the omissions (reported after pictures) in nonvisible response categories, 78% were classified as reminded, 13% neglected, and 9% forgotten.

What proportion of participants' *total* remindings were nonvisible? Collapsing across testing sessions and averaged across participants, the mean proportion of remindings that were in the nonvisible response categories was .22 ($SD = .14$, 95% CI [.11, .34]), with the individual participant proportions being: .00, .09, .11, .18, .22, .27, .30, .35, and .50. Eight out of our 9 participants experienced at

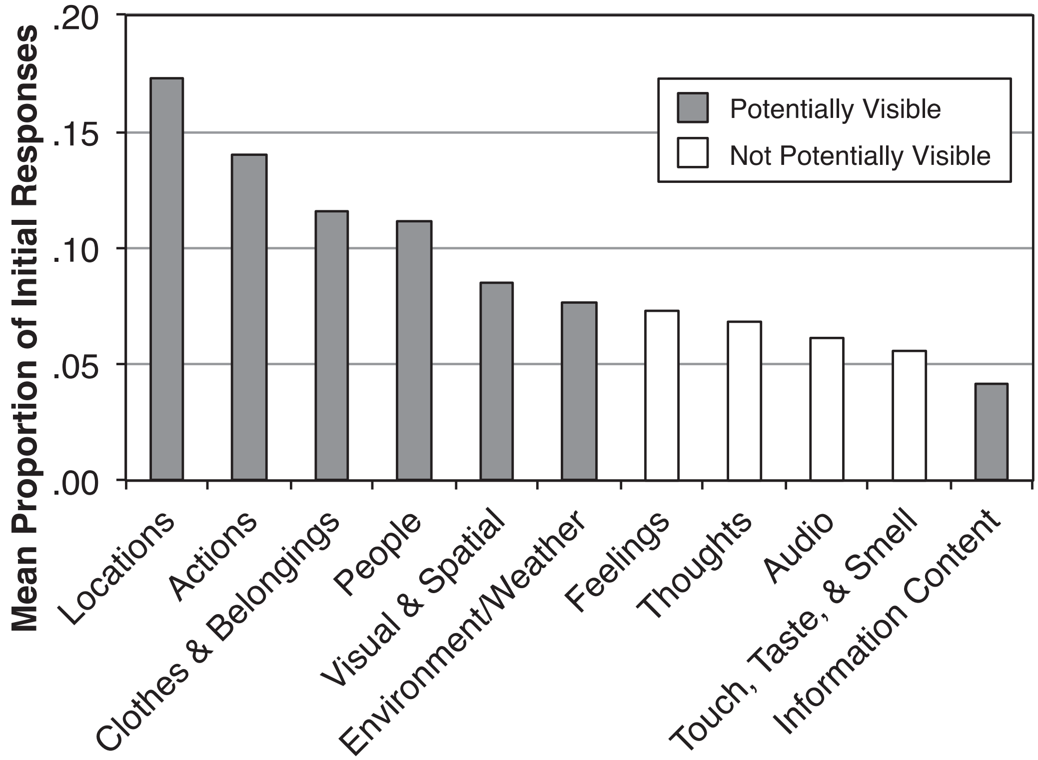

Figure 3. Mean proportion of initial recall responses per response category.
Note: Collapsed across testing sessions and averaged across participants.

least one non-visual reminding upon being cued with their SenseCam pictures. The mean (and *SD*) of the proportions for each of the nonvisible categories were: .09 (.15) for thoughts; .05 (.03) for touch/taste/smell; .03 (.03) for audio; and .02 (.03) for feelings. If we limit consideration to just thoughts and feelings, the mean proportion of remindings was .12 ($SD = .16$), with 5 out of our 9 participants experiencing such remindings.

The mean (and *SD*) of the ratings for nonvisible remindings were as follows: 6.64 (0.47) for confidence; 4.88 (1.14) for reliving; 4.29 (1.34) for knowledge; and 3.86 (1.41) for frequency. Compared to the same data for the visible categories (Table 3), the nonvisible remindings show the same overall pattern with the exception of higher knowledge ratings. Here we provide some examples of remindings from nonvisible categories. Thoughts: "I remember thinking that it was strange that the man and woman riding in the elevator with us existed outside of IGB and had their own lives and such.". Audio: "The strange noise it makes on the Blockbuster Express touchscreen machine". Touch/taste/smell: "I also kept on drinking the soday. I twas sweet.". Feelings: "Happiness and fulfillment.". A number of the remindings from the information content category could also be considered nonvisible. For example: "We started randomly listing off words that rhymed with 'near' until we couldn't think of anymore".

These nonvisible remindings corroborate anecdotal and empirical reports that watching a movie of one's SenseCam pictures powerfully triggers a sense of reliving the original experience, even including aspects not visible in the pictures themselves. Loveday and Conway (2011b) argued that when pictures are used as retrieval cues they often lead to "Proustian moments" in which "images of the past flood into consciousness and the rememberer has a powerful experience of recollection" (p. 697).[6] Consistent with this claim, the amnesic patient

Table 2. Repeated measures correlations among initial recall response ratings.

	Confidence	Reliving	Knowledge	Frequency	Accuracy
Confidence	–				
Reliving	**.35**	–			
Knowledge	**-.07**	**-.55**	–		
Frequency	.01	**-.31**	**.36**	–	
Accuracy	**.26**	*.05*	.01	**.06**	–

Note: Data are only from potentially visible response categories, collapsed across retention intervals; rating measures were all made on 1–7 scales; accuracy was self-scored after seeing pictures; $N = 9$; $df = 1{,}075$; **boldface**: $p < .05$; *italics*: $p < .10$; see Bakdash and Marusich (2017) for specification of repeated measures correlation; in Supplemental Materials see Table S4 for 95% confidence intervals, and Table S5 for mean Pearson correlations, showing the same pattern.

Table 3. Means (and standard deviations) of ratings for accurate initial responses and omissions (reported after pictures).

Response type	Response count	*n*	Ratings at time of response				Response length
			Confidence	Reliving	Knowledge	Frequency	
Initial: Accurate	922	9	6.3 (0.5)	4.3 (1.0)	4.2 (1.3)	4.4 (0.9)	7.2 (4.1)
Omissions (Reported After Pictures)							
Neglected	27	6	6.8 (0.3)	5.3 (1.1)	4.3 (1.4)	5.1 (1.5)	4.0 (1.8)
Reminded	203	9	6.8 (0.2)	4.4 (1.1)	3.8 (1.3)	3.3 (1.0)	11.1 (8.9)
Forgotten	60	6	6.7 (0.7)	2.1 (0.7)	5.3 (1.3)	5.1 (1.1)	5.2 (2.3)

Note: Data are only from potentially visible response categories, means of participant means, collapsed across two retention intervals (approximately 1 week and 1 month); *n* is the number of participants (out of 9) who gave each type of response; rating measures were all made on 1–7 scales; Initial: Accurate responses were those self-scored by participants as 7 out of 7 on accuracy after seeing pictures; Response Length is number of words.

in their study was able to recall significantly more nonvisible details (e.g., thoughts, feelings) when cued with her SenseCam pictures versus a written diary. Loveday and Conway also recount an anecdote of a colleague who was watching his pictures of a walk to a meeting three months prior and suddenly remembered the song he had been listening to on his iPod at that moment. Similar anecdotes of non-visual remindings were reported in a poster by Berry et al. (2006). For example: "My feet were damp and cold", "I remember thinking that [T] was quiet and shy", and "Talking to a barmaid about her giving piano lessons".

We know that even verbal retrieval cues can elicit memory of non-visual aspects of experience, as shown by the initial recall data in Figure 3. Similarly, Brewer (1988) found that verbal descriptions of the actions in an event elicited recall about thoughts in 42% of the cases. Our current data confirm that pictures from the original experience can elicit additional non-visual recall, beyond what could be recalled from verbal cues alone.

What are the theoretical implications of these descriptive data? Brewer (1996, p. 38) pointed out an inconsistency in his earlier (1986) analysis of recollective memory. Brewer (1986, Table 3.1) implied that recollective memory is represented in image form, where "image" refers to the mental representation of any sensory experience, not just visual; yet he also stated that there are non-image components to a recollective memory experience (p. 34). Brewer's (1986) definition stated that recollective memory "frequently appears to be a 'reliving' of the individual's phenomenal experience during that earlier moment" (p. 34). It seems to us that the resolution to these complex issues is to assume that the structure of recollective memory and the experience of "reliving" do not just include the original ongoing sensory input, but *all* of the original ongoing mental processes, including thoughts and feelings (see Brewer & Pani, 1983, p. 15 for a discussion of "phenomenal states that are nonimaginal"). This account predicts the phenomenon of Proustian moments: when one aspect of original experience is reinstated (e.g., the visual aspect, as cued by watching a movie of one's SenseCam pictures), then by the action of spreading activation one may experience a multi-modal sense of reliving, extending to other aspects of the original experience including non-visual sensation, thoughts, and/or feelings (see also "redintegration"). An ongoing line of the first author's research seeks to explore such non-visual remindings (Finley et al., 2023).

Completeness of autobiographical memory

In this section, we examine the issue of *completeness* of autobiographical memory: what proportion of their original experience did participants recall? In order to study this issue one needs a measure consisting of the amount of information recalled divided by the total amount of information experienced. The challenge with such a measure for autobiographical memory is that we must decide how to quantify the participants' original experiences in order to form the denominator. We discuss this thorny issue in detail in the Supplemental Materials, where we justify the practical formula we arrived at.

The numerator in our formula for completeness consists of the count of initial recall responses (i.e., those made before seeing a timeslice's pictures) plus omissions classified as *neglected* (i.e., those that participants stated they had remembered before seeing the pictures but had simply neglected to report). The denominator is the count of initial recall responses plus *all* types of omissions (*neglected*, *reminded*, and *forgotten*). We then need a method to handle the counting of *erroneous* initial recall responses. We have chosen to dichotomise the initial recall responses into accurate (self-scored 7 on the scale of 1–7) versus inaccurate (self-scored 6 or lower), and only include accurate responses in our formula. By analogy, in a laboratory free recall experiment, one would not include intrusions (i.e., recalled words that were not on the list) when calculating the proportion of the original list that a participant recalled. Thus, our formula for completeness is:

$$completeness = \frac{(accurate\ initial\ recalls + neglected\ omissions)}{(accurate\ initial\ recalls + all\ omissions)} \quad (1)$$

Thus, the question we are asking is: how much were participants able to remember, unaided, about their original experiences compared to how much they were able to report about those experiences when provided with a pictorial record? We understand that this approach yields a measure of completeness that is too large, because the total amount that participants can report with the help of the pictures must be considerably smaller than the totality of their original experience. Nevertheless we think it is important to begin providing data on this issue, even if all we can do is suggest an *upper bound estimate* of the completeness of autobiographical memory for everyday experiences.

Using this formula we calculated overall completeness for each participant, collapsing across timeslices and testing sessions, for the visible response categories only. The mean proportion completeness across participants was .79 (*SD* = .11, 95% CI [.70, .88]), with the individual participant proportions being: .63, .66, .70, .75, .80, .80, .86, .89, and .99.[7] We have conceived of these data as estimating an upper bound on completeness; we can also conceive of them as estimating a lower bound on forgetting (i.e., 1 – completeness). The mean proportion forgotten was .21 (*SD* = .11, 95% CI [.12, .30]), with the individual participant proportions being: .01, .11, .14, .20, .20, .25, .30, .34, and .37.

So the data suggest that our participants were able to recall *at most* about 79% of their original experiences from a week to a month earlier, and had forgotten at least about 21%. It would clearly be worthwhile to explore such data at shorter and longer retention intervals. We understand that we have not solved the difficult problem of the amount of information *experienced* in everyday life (see Supplemental Materials). However, we have at least made some initial steps by using our data to establish an upper bound on how much of everyday experience is remembered.

Comparable retention curves in autobiographical memory

How do the completeness data from our study compare to data from previous studies of autobiographical memory? As we discussed in the Introduction, studies of autobiographical memory have used a variety of methods. We have selected the previous studies that used planned records (daily diary, experience sampling, and wearable cameras) that yielded completeness data most compatible with those of the current study (Brewer, 1988; Finley et al., 2011; Linton, 1975, 1978, 1982; Thompson et al., 1996; Wagenaar, 1986; White, 1982, 1989, 2002). A comparison of the data from those studies with our own is presented in Figure 4, with retention intervals out to 2 years. Data from the current study were from the visible response categories only, averaged across participants separately for the two testing sessions. A brief description of the methods and measures used in each of the previous studies is given in the Supplemental Materials. To our knowledge, this is the first time these disparate data have been brought together.

Figure 4 shows that overall performance levels are remarkably high, yet do show a steady decline. One noteworthy pattern is that the different levels of performance across studies appear to correspond to the method that each study used to sample subsets of original experience for testing. *The more random the sampling method, the lower the overall performance.* Linton and Wagenaar made written records of *distinctive* events from each day, and their performance was substantially higher than that of White, who did not have a consistent strategy for selecting events. The participants in the diary studies by Thompson et al. were told to select a variety of events, both memorable and not memorable, while also keeping events unique. Their rating performance at the shorter intervals began at the ceiling like Linton and Wagenaar's data, while their location recall performance at the longer intervals fell below Linton and Wagenaar's, but above White's. The latter pattern likely reflects the mixture of both memorable and not memorable events. The practical reason for selecting distinctive events in all these studies was to make testing unambiguous: each written record had to uniquely cue memory back to that specific event and no others. So common everyday events, such as brushing one's teeth, could not be used. Such common event cues could also be answered with semantic knowledge of one's life rather than a specific episodic recollection. However, distinctive events are also more memorable (Hunt & Reed, 2006), and the high performance levels of Linton's and Wagenaar's studies reflect this.

Brewer's, 1988 study allows an explicit examination of the effect of selecting *memorable* events on autobiographical memory performance. In that study, the core data were based on records of events that were randomly selected (cued by a beeper), but the participants were also asked, at the end of each day, to make a record for "the most memorable event that occurred that day". As shown in Figure 4, performance on the memorable events was higher than on the randomly selected events. This finding suggests that the participants showed successful metamemory judgment in selecting everyday events that would yield high recall. Brewer went on to attempt to discover the characteristics that the participants had used to successfully select more memorable events. Table 3.1 (p. 36) of Brewer's, 1988 study compared the characteristics of memorable events with random events and showed that the participants were likely relying on a variety of event characteristics, but the event attribute that was the strongest was *Low frequency* of occurrence. The memory data in that study validated the participants' metamemory beliefs about low frequency of occurrence leading to better memory. In the current study, we also found that lower frequency events were accompanied by higher ratings of reliving ($r_{rm}(1075) = -.31$, Table 2).

As original experience is sampled more randomly, it becomes more likely that participants will be tested on the mundane and trivial moments that make up the fabric of everyday life. It should not be surprising if such fabric is poorly remembered compared to meaningful or distinctive experiences. Three studies plotted in Figure 4 used truly random sampling of events (the current study; Brewer, 1988, random; and Finley et al., 2011), and they

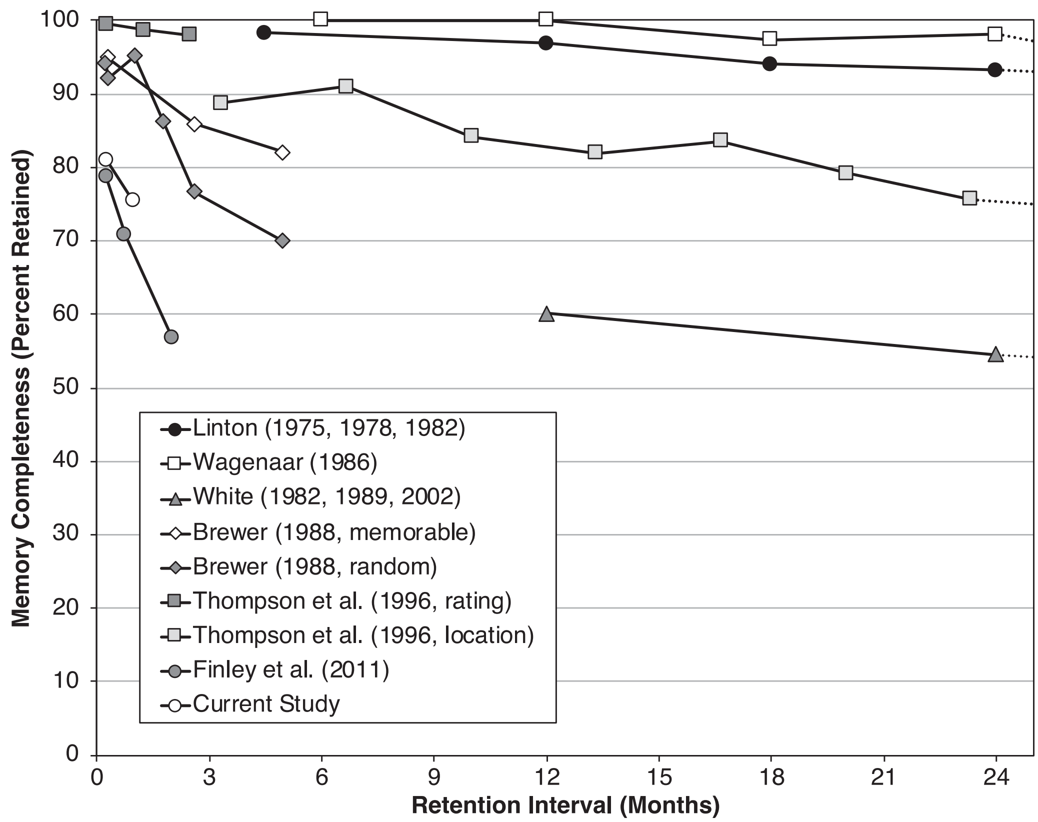

Figure 4. Comparison of completeness results across several autobiographical memory studies.
Note: See Supplemental Materials for details on methods and measures for each study.

showed markedly lower performance than the studies by Linton and Wagenaar which used distinctive events, and the studies by Thompson et al. which used a variety of unique events. Interestingly, the curve from White's study, which used a vaguely defined "haphazard" selection of events, looks almost like an extension of the random curve from Brewer's study. The two studies that used wearable cameras (the current study, and Finley et al., 2011) went even further than previous studies in that we did not just make records of individual discrete events, but in fact made a continuous record of an entire day's experience, which we then tested for in segments. These segments necessarily contained some memorable experiences as well as many forgettable ones. But Figure 4 also shows that memory performance for randomly selected events was higher in Brewer's, 1988 study than it was in the two SenseCam studies. A key methodological difference that may account for this is that recording was passive in the SenseCam studies. By contrast, in Brewer's study participants actively wrote down their experience in narrative form when prompted, which may have enhanced memory (p. 82; MacLeod & Bodner, 2017; Roediger & Butler, 2011). Nevertheless, even with the continuous passive recording used in the current study, the completeness of participants' memory for everyday experience was still impressively high after one week to one month. Also noteworthy is the similarity in completeness between the current study and our previous SenseCam study (Finley et al., 2011), considering that the previous study used a recognition rating (1–7) while the current study used the more complex measure of completeness detailed earlier.

Finally, how can we explain the fact that many of these retention curves show a much slower and steadier decline than the classic Ebbinghaus forgetting curve which features rapid initial decline followed by deceleration? One explanation is simply that experiences from everyday life are more meaningful than typical laboratory stimuli, and thus memory performance for them declines more slowly. In our upcoming discussion of episodic versus semantic autobiographical memory, we note that participants can produce accurate responses based on recollection and/or general knowledge of their own lives. For meaningful stimuli such as the distinctive events used in the studies with the highest curves, as memory retrieval shifts from reproduction to reconstruction with increasing retention intervals, participants can use the familiar schemas of their lives to maintain high performance even as recollection declines. This strategy becomes less effective for more random slices of life, especially ones cued with indiscriminately captured pictures.

Notice that for the three truly randomly sampled studies (current study; Brewer, 1988, random; Finley et al., 2011), not only are the curves lower than the others, but they are also the only ones that show the steep initial decline characteristic of the Ebbinghaus forgetting curve. Thus, across the wide range of methods used to study the rate of forgetting in autobiographical memory, the pattern is that memory for events selected as distinctive or memorable shows extraordinarily little forgetting over years, while memory for events selected randomly shows more and faster forgetting but still with impressive levels of retention after months.

Memory increase due to pictures

The very existence of the wearable camera used to conduct this research is a hallmark of the increasing role that external memory stores are playing in human lives (e.g., mobile camera phones, social media). In 1919, Marcel Proust stated that "the greater part of our memory exists outside us" in the form of cues that can revive the otherwise irrecoverable past (Proust 1919/2002). In 1945, Vannevar Bush conceived of a then-theoretical recording and storage device, *memex*, that would provide an individual with "an enlarged intimate supplement to his memory". In 1995, Schönpflug and Esser argued that "individual memory should be analyzed as a component of an extended system that also comprises social and technical supports" (see also Finley et al., 2018; Nestojko et al., 2013; Schönpflug, 1986). Advanced external supplements to human memory have become a reality with the development of personal memory technologies such as wearable cameras, and with the advent of lifelogging: automatic recording of many aspects of everyday experience (Bell & Gemmell, 2009; Dobbins et al., 2014; Kalnikaite et al., 2010; van den Hoven et al., 2012; Vemuri, 2005).

Pictures have tremendous potential as external memory aids (Fawns, 2020; Foley, 2020; Henkel et al., 2020; Soares, 2023). Viewing pictures taken by SenseCam has helped memory-impaired individuals to better recall everyday experiences (Berry et al., 2007; Doherty et al., 2013; special issue of *Memory* edited by Loveday & Conway, 2011a). Our data allow us to consider, for healthy participants with unimpaired memory, how much more they were able to remember by looking at their pictures compared to their initial attempts to remember without their pictures. For each participant, collapsing across retention intervals and for the potentially visible response categories only, we calculated a proportion: the numerator was the number of omissions (reported after pictures) classified as *reminded*, and the denominator was the number of initial recall responses self-scored as totally accurate plus the number of omissions classified as *neglected*.

$$\textit{memory increase} = \frac{\textit{reminded omissions}}{(\textit{accurate initial recalls} + \textit{neglected omissions})} \quad (2)$$

This calculation tells us how much more the pictures helped participants remember relative to what they had already remembered without the pictures. The mean proportional increase in memory across participants was .23 ($SD = .14$, 95% CI [.11, .34]), with the individual participant proportional increases being: .01, .12, .16, .16, .18, .21, .30, .42, and .46. A within-subjects *t*-test conducted on the number of responses (reminded + accurate + neglected vs. accurate + neglected) showed this was a statistically significant increase, $t(8) = 4.05$, $p = .004$, $d = 0.30$.

External memory is valuable not only for stimulating internal memory, but also for providing objective records of that which cannot be remembered. Thus, we also consider how much additional *information* participants gained by looking at their pictures, regardless of whether or not they remembered that information. For this calculation, we simply added the omissions classified as *forgotten* in the numerator of the previous formula.

$$\textit{information increase} = \frac{(\textit{reminded omissions} + \textit{forgotten omissions})}{(\textit{accurate initial recalls} + \textit{neglected omissions})} \quad (3)$$

The mean proportional increase in information across participants was .29 ($SD = .18$, 95% CI [.15, .44]), with the individual participant proportional increases being: .01, .12, .16, .24, .24, .33, .43, .52, and .59. A within-subjects *t*-test conducted on the number of responses (reminded + forgotten + accurate + neglected vs. accurate + neglected) showed this was a statistically significant increase, $t(8) = 3.80$, $p = .005$, $d = 0.37$.

So the data suggest that, after a week to a month, looking at pictures of original experiences provided about a 23% increase in memory and a 29% increase in information. Pictures even reminded participants of nonvisual aspects of experience, as we discussed earlier in the section on Nonvisible Remindings. These findings clearly illustrate the reminding power of pictures from one's everyday life.

Overall timeslice ratings. Did participants think that seeing the pictures increased their memory? Yes. Recall that participants made several *overall timeslice ratings* on a 1–7 scale, once just after the initial recall phase (before pictures), and once just after the self-scoring phase (after pictures). Their memory strength ratings increased from before pictures (mean of participant means = 3.81, SD = 0.91) to after pictures ($M = 4.75$, $SD = 0.89$), $t(8) = 2.82$, $p = .022$, $d = 0.94$. Their reliving ratings also increased from before pictures ($M = 3.31$, $SD = 0.80$) to after pictures ($M = 4.26$, $SD = 1.00$), $t(8) = 2.58$, $p = .033$, $d = 0.86$. After the self-scoring phase, participants rated how much more they remembered of the timeslice once they had seen the pictures. Their subjective ratings ($M = 5.34$, $SD = 0.71$) correlated positively with the objective memory increase measure (Formula 2) calculated at the level of timeslices, $r_{rm}(112) = .32$, 95% CI [.14, .47], $p < .001$. Thus participants showed an understanding of the reminding power of pictures.

Accuracy of autobiographical memory

In this section, we examine the issue of the *accuracy* of everyday autobiographical memory: how much of what participants remembered actually happened as they remembered it? We will examine this largely using the rate and nature of errors.

Error rate and accuracy rate. Collapsing across testing sessions and participants, and excluding miscued time-slices, there was a grand total of 1,460 initial recall responses, 1,085 of which were in the visible response categories. Of those 1,085 responses in the visible categories, 163 were self-scored as less than completely accurate (i.e., rated 1–6 out of 7 on the accuracy rating scale). Of those 163, we excluded 62 from accuracy and error analyses for several reasons: 33 because participants did not follow instructions (e.g., writing "n/a" instead of text to correct their initial response), 27 because participants wrote that the pictures did not give enough information to score their initial response, and 2 because they wrote something we could not interpret.[8] Thus there were 1,023 responses (1,085–62 excluded) that we were able to unambiguously classify, and 101 of them were self-scored by the participants as clear memory errors. The mean number of such errors per participant, collapsing across testing sessions and considering only classifiable responses in the visible categories, was 11.22 (*SD* = 5.57, 95% CI [6.68, 15.77]), with the individual participant values being: 1, 6, 8, 9, 12, 14, 15, 15, and 21. Thus, all participants made at least one clear error. The mean *proportion* of errors per participant was .11 (*SD* = .06, 95% CI [.06, .17]), with the individual participant proportions being: .05, .06, .06, .09, .10, .11, .15, .15, and .26.

How much of participants' recall of their day was accurate? The accuracy rate is simply the inverse of the error rate. The mean proportion accurate was .89 (*SD* = .06, 95% CI [.83, .94]), with individual participant proportions being: .74, .85, .85, .89, .90, .91, .94, .94, and .95. What about the accuracy ratings? Across participants the mean accuracy rating was quite high at 6.52 (*SD* = 0.26, 1–7 scale, 95% CI [6.31, 6.74]), with the individual participant means being: 6.06, 6.23, 6.33, 6.47, 6.48, 6.71, 6.78, 6.79, and 6.88.

How reliable is our finding of an 11% error rate? One way to check is by making a comparison to data from Brewer's 1988 beeper study of autobiographical memory. We re-analysed the data from that study (Table 3.8) that most closely corresponded to the type of data in the current study (as detailed in Supplemental Materials), and found an error rate of 10.67%, which is very close to the error rate of 11.46% in the current study. Given the many methodological differences between the two studies, we cannot know how much of the similarity in error rate may be due to chance. Nevertheless, the findings together give us some confidence that the order of magnitude for the error rate in everyday autobiographical memory is around 10% at retention intervals ranging from a week to one or two months.

What are the theoretical implications of these data? On the one hand, an error rate of approximately 11% provides further convincing evidence against strong copy theories of memory (e.g., Brown & Kulik, 1977; Penfield, 1969). On the other hand, the great majority of recalls (89%) were reasonably accurate descriptions of the everyday events that had occurred one week or one month earlier in the lives of these college undergraduates. Therefore the data also seem to weigh against strong reconstructive theories of memory (e.g., Barclay, 1986; Neisser, 1982, pp. 43–48). Instead, the data seem quite consistent with Brewer's (1986) *partially reconstructive* view which stated that recollective memories (personal memories) are "reasonably accurate copies of the individual's original phenomenal experience" (p. 43) but that they are subject to a variety of reconstructive processes that can cause errors, just as are other forms of memory (see reviews in Roediger & DeSoto, 2015; Schacter, 1995).

Classes of errors. On the basis of previous research on errors in autobiographical memory (e.g., Brewer, 1988; Hyman et al., 1998), combined with a lengthy examination of our participants' inaccurate recalls and corresponding corrections, we developed a set of five classes that accommodated the errors in our data, shown in Table 4. Both authors first independently classified all 101 memory errors into one of the five classes (*time shift, substitution, intrusion*, *distortion,* and *false assertion of absence*). We then resolved any discrepancies through discussion. The Appendix gives additional examples from each of the five error classes to provide the reader with a qualitative feel for errors that occur in everyday autobiographical memory. Table 5 shows the *confidence*, *reliving*, *knowledge*, and *frequency* ratings for the five classes of errors as compared to the accurate initial responses.

Memory processes leading to errors. Why did participants make the errors that they did? We think it is likely that a number of established memory processes were at work. For example, most of the *time shift* errors are probably best thought of as retrieval errors (see *time slice*, Brewer, 1992, pp. 290–291; Hyman et al., 1998). When one of our participants recalled, "I practiced my song in the car", and then after viewing the evidence in the pictures stated, "I did this the next week", we think it is likely that on some other day in the past (for which we have no picture evidence) she did in fact sing in the car and retrieved this past episode by mistake when trying to recall the particular day we asked about.

Some of the *substitutions* and *time shift* errors suggest the impact of interference. For example, one of our participants recalled having her "Black messenger bag" with her, and then after seeing the pictures said, "I actually brought my white and black backpack instead of my messenger

Table 4. Classification of memory errors.

Error class	Response count	Description	Example initial response	Example of participant's correction notes
Time Shift	42	Participant stated that the response occurred at another time.	[Actions] *Prepared my bag for the day.*	*I didn't do this till later*
Substitution	23	Response substituted one entity for another.	[Clothes & Belongings] *I was wearing my England football jersery*	*I was wearing a soccer jersey but not my England on[e] it was my Mexico club soccer one.*
Intrusion	20	Response stated something that did not occur at all.	[Actions] *layed in bed*	*thought I did, guess i didn't*
Distortion	11	Response distorted some characteristic of an entity.	[Visual & Spatial] *We were playing on [N]'s round table.*	*The table has angles.*
False Assertion of Absence	5	Response falsely asserted absence of some entity.	[People] *I was alone.*	*I wasn't alone.*

Note: This classification scheme was applied to initial recall responses self-scored as inaccurate, for the visible response categories only; response count is across all 9 participants. Response categories are shown in brackets. Names of people are replaced with bracketed initials to preserve privacy.

bag.". We think it is quite likely that these two functionally similar objects caused interference when she had to recall which one she had on a particular day. Similarly, a different participant said "Backpack" then after seeing the pictures said, "had purse, not backpack". Other substitutions that were likely due to interference included: one friend confused for another, one parking lot confused for another, walking home versus driving home, drinking ginger ale versus Sprite soda, wearing a t-shirt versus a sweatshirt, and hearing the sounds of a video game versus a TV broadcast of a basketball game.

On the day before the acquisition period, participants had completed a form detailing their anticipated schedule for the upcoming days. This form gave us some insight into participants' daily scripts (i.e., their semantic knowledge of their typical daily routines). This helped us to observe that a number of *substitution* and *intrusion* errors likely reflected the process of semantic information overriding a less frequent episodic event memory. For example, one of our participants recalled that he "flipped through the channels on the TV, watching things" (confidence 5, reliving 2, knowledge 6, and frequency 6) and then after viewing the pictures stated that he had "actually watched the movie Wolfman.". This timeslice occurred in the evening between 10 and 11pm. On the anticipated schedule form for that day, the evening entries included "watch tv or read" at 7 pm and "maybe drink w/ friends, otherwise watch movie" at 9 pm. This activity pattern was the same across 3 of the 5 weekdays in the anticipated schedule. It seems plausible to us that when the participant was trying to recall the particular timeslice he was unable to remember the episodic event of watching Wolfman and thus substituted semantic information from his daily TV watching script, writing "watching things" as a catch-all to cover TV shows and movies. This explanation is consistent with the initial recall ratings, which were low in reliving (lack of specific episodic memory) and high in both knowledge and frequency (relying on a typical evening script).

The *false assertion of absence* errors were the rarest type of error, and the only type that we did not anticipate. For example, when a participant initially recalled a specific person being absent from an event, but then saw from the photos that the person was in fact present. These errors featured high confidence, reliving, and knowledge ratings. We suspect they were based on participants' *schemas* for their everyday life. If their schema holds that a certain entity should be present in a given situation, and they can recall other aspects of that situation but not the entity, that absence would stand out as a schema violation (cf. Brewer & Treyens, 1981, p. 224). In contrast, failure to recall the same entity in some other situation (where it is not expected) would not stand out as something worth mentioning.

Reconstructive imagery. Some errors in human memory may be experienced in image form, as suggested by Brewer and Pani (1983). Brewer (1986) argued that "the

Table 5. Means (and standard deviations) of ratings for accurate and error initial responses.

			Ratings at time of recall (Before pictures)				
Response type	Response count	*n*	Confidence	Reliving	Knowledge	Frequency	Accuracy rating (After pictures)
Accurate Responses	922	9	6.3 (0.5)	4.3 (1.0)	4.2 (1.3)	4.4 (0.9)	7.0 (0.0)
All Errors	101	9	5.4 (1.0)	4.1 (1.4)	4.4 (1.3)	4.5 (0.9)	2.9 (1.1)
False Assertion of Absence	5	3	6.5 (0.4)	4.7 (1.2)	5.2 (1.4)	4.0 (1.9)	4.8 (0.6)
Time Shift	42	7	6.0 (1.1)	3.9 (1.8)	4.5 (1.7)	3.6 (1.1)	2.6 (1.3)
Substitution	23	9	5.8 (1.2)	4.2 (1.4)	4.2 (1.4)	4.7 (1.0)	2.9 (1.1)
Distortion	11	7	5.6 (1.4)	4.0 (1.5)	3.4 (1.7)	4.6 (1.6)	4.5 (0.5)
Intrusion	20	6	4.2 (1.3)	2.6 (1.6)	5.1 (1.1)	5.8 (1.1)	1.5 (0.9)

Note: Data are only from potentially visible response categories; values are means of participant means, collapsed across retention intervals; *n* is the number of participants (out of 9) who gave that type of response; rating measures were all made on 1–7 scales; error rows are sorted in descending order by confidence ratings.

original experience can be reconstructed to produce a new nonveridical [recollective memory] that retains most of the phenomenal characteristics of other [recollective memories] (e.g., strong visual imagery, strong belief value)." (p. 44). Consistent with this claim, the data in Table 5 show that the reliving ratings for the erroneous responses ($M = 4.1$) were not significantly different from the reliving ratings for the accurate responses ($M = 4.3$), Wilcoxon matched-pairs signed-rank test, $z = 0.65$, $p = .570$. Averaged across participants, 32% of the self-scored-as-accurate responses were given a rating at the top of the reliving scale (6 or 7), and 37% of the errors were given a similar rating (6 or 7). Rubin et al. (2003) showed that a question about *reliving* is essentially equivalent to a question about *mental imagery*. Our data thus appear to provide clear direct evidence that many errors in autobiographical memory are experienced in the form of reconstructive visual imagery.

Overall, memory for recent everyday experiences appears to be mostly accurate, while showing some systematic types of errors, consistent with Brewer's partially reconstructive view of autobiographical memory performance (1986, p. 43).

Episodic versus semantic autobiographical memory

Each of the recalls of timeslice details (response rows) was rated by participants on the degree to which it was experienced as a *reliving* of the original event and also on the degree to which it was based on the participants' general *knowledge* of their lives. We used these data to explore the episodic/semantic distinction in autobiographical memory for everyday experiences by seeing if participants classified some responses as predominately episodic and others as predominately semantic. We classified a response as episodic if its reliving rating was higher than its knowledge rating, and we classified a response as semantic if its knowledge rating was higher than its reliving rating. Figure 5 provides the descriptive data on this issue. The figure shows, averaged across testing sessions and then across participants, the number (represented as bubble diameter) of initial recall responses for all combinations of knowledge and reliving ratings for the potentially visible response categories only. The data displayed in the figure support the distinction between semantic and episodic autobiographical memories. There is a tendency for one set of responses to cluster toward the top-left corner of the figure and another set to cluster toward the bottom-right corner of the figure. Responses with equal reliving and knowledge ratings (i.e., those few along the diagonal) were excluded from all further episodic/semantic analyses. Across participants, the mean number of episodic responses per test was 32 ($SD = 28$), and the mean number of semantic responses per test was 23 ($SD = 15$). These did not significantly differ, $t(8) = 0.77$, $p = .462$, $d = 0.35$.

Characteristics of episodic and semantic autobiographical memory

Frequency. Brewer (1986, 1996) has argued that exposure to single distinctive events tends to produce strong recollective memories (episodic). This claim would predict that the responses we classified as episodic should be rated as lower in frequency than those we classified as semantic. The mean frequency rating for the episodic responses ($M = 3.6$, $SD = 0.8$) was indeed significantly lower than that for the semantic responses ($M = 4.9$, $SD = 0.9$), $t(8) = 3.89$, $p = .005$, $d = 1.39$. Furthermore, the correlation between frequency and episodic/semantic score (reliving minus knowledge rating) was negative, $r_{rm}(1075) = -.38$, 95% CI [−.42, −.32] $p < .001$. Thus, the data support Brewer's characterisation of episodic recollective memories.

Accuracy. In Figure 5, the shade of the bubbles represents the mean accuracy of responses with that combination of reliving and knowledge ratings, with darker shades representing higher accuracy. We reason that for many aspects of autobiographical memory one can arrive at a correct answer to a question by using episodic *and/or* semantic information (see Brewer, 1986, pp. 25–26; Brewer & Pani, 1983, pp. 2–4, 31–32). Thus if asked what you had for breakfast yesterday you could accurately answer the question by recollecting this morning's colourful bowl of Sprinkle Spangle cereal with fried eggs and kale (episodic memory), *and/or* by knowing that you eat Sprinkle Spangle with fried eggs and kale every morning (semantic memory). Given these dual sources of information about autobiographical memory, we see no way to make any a priori claims about the accuracy of episodic and semantic memory and think we should let the data speak to the issue. In our data, there was no difference in accuracy between episodic and semantic autobiographical memory. The mean accuracy rating for episodic responses (6.4, $SD = 0.5$) did not reliably differ from that for semantic responses (6.3, $SD = 0.3$), $t(8) = 0.89$, $p = .402$, $d = 0.27$. Our data thus suggest that for a sample of everyday autobiographical memories of undergraduate life, the responses we classified as more episodic and those we classified as more semantic were of roughly equivalent accuracy. This can be seen visually in Figure 5 by noting that neither of the triangular sections is obviously darker overall than the other. The key point here is that *for autobiographical memory there are two different ways that participants can produce an accurate response: through episodic recollection, and through semantic knowledge of their own lives.*[9]

Confidence. We know of no theory that makes predictions about the relative level of confidence shown for episodic versus semantic memory, although some relevant data do exist. In laboratory list-learning experiments using recognition tests with *remember/know* judgments, participants generally give higher confidence ratings to

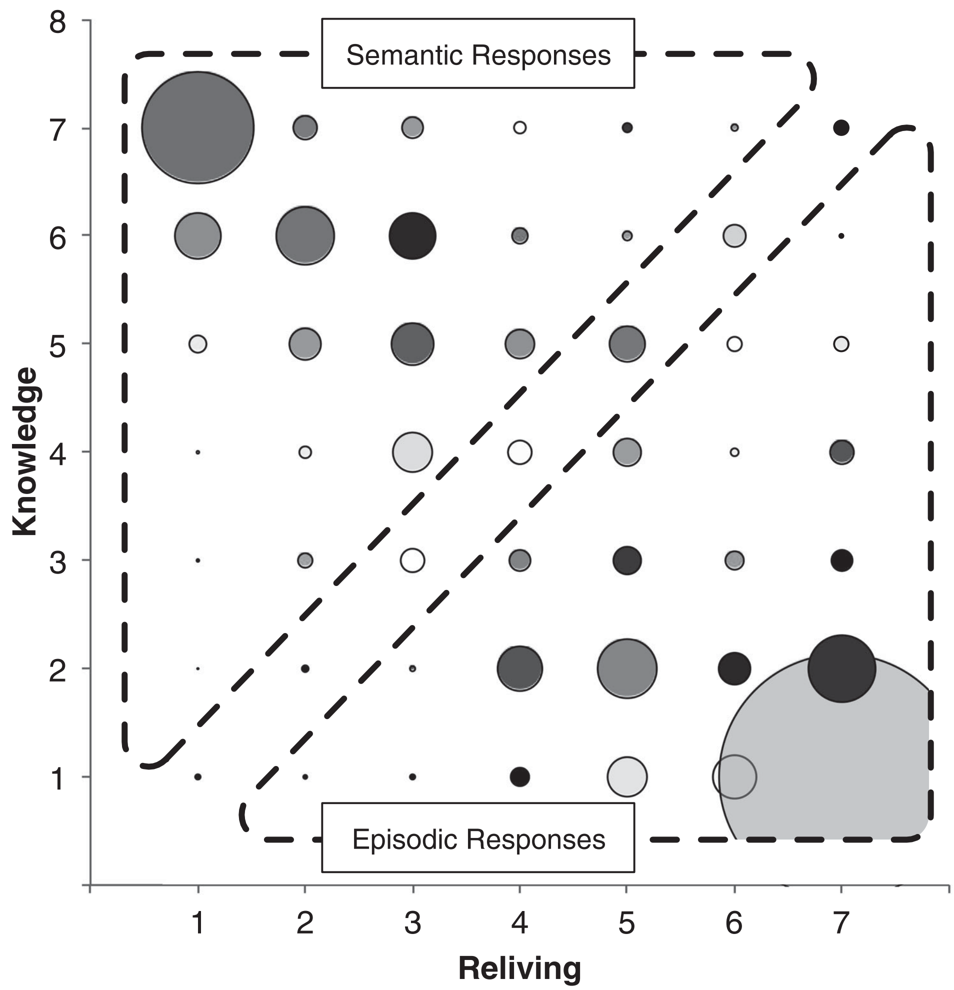

Figure 5. Bubble plot illustrating semantic versus episodic responses.

Note: Number (bubble diameter) of initial recall responses for all combinations of knowledge and reliving ratings, collapsed across testing sessions and averaged across participants, for the potentially visible response categories only. Darker shades represent higher mean accuracy ratings.

responses based on recollection versus responses based on familiarity (Wixted & Mickes, 2010); however, we do not think that the *remember/know* distinction is a good reflection of the *episodic/semantic* distinction (see Brewer & Lockhart, 1995; Brewer & Sampaio, 2006; Kihlstrom et al., 1996). In the domain of autobiographical memory, there are a number of studies that show that recollective imagery, the hallmark of autobiographical episodic memory, correlates positively with confidence (Table 2; Brewer, 1988, Table 3.14; Talarico & Rubin, 2003, 2007). In addition, previous research has shown that one can find very high levels of confidence for episodic memory tasks (e.g., Brewer & Sampaio, 2006) and for semantic memory tasks (e.g., Brewer & Sampaio, 2012). Consistent with these sets of empirical findings across a variety of domains, our data for autobiographical memory show that both episodic (6.4, $SD = 0.6$) and semantic responses (5.7, SD = 0.4) demonstrated high levels of confidence on the 1–7 scale, with confidence being higher for episodic responses, $t(8) = 3.68$, $p = .006$, $d = 1.15$. There were also interesting differences in response categories and response length between episodic and semantic responses, covered in Supplemental Materials along with other details in Table S6.

Revisiting the episodic/semantic distinction

The data presented in Figure 5 certainly suggest that our participants used the reliving and knowledge rating scales in ways that are consistent with the episodic/semantic distinction. However, this is a very complex issue. Conway has theorised that memories for specific episodes are embedded in a broader hierarchical framework of autobiographical knowledge (Conway & Pleydell-Pearce, 2000), and furthermore that the episodic memories themselves consist of both episodic elements (e.g., a visual representation of a moment) and an organising conceptual knowledge frame (Conway, 2009).

We take a position similar to that of Conway and propose that recollective memories for everyday experiences can be composed of pieces of old knowledge (semantic) configured into a new pattern (episodic) determined by the unique properties of the particular experience being remembered. This makes theoretical understanding of the episodic/semantic distinction difficult in the domain of autobiographical memory. For example, instead of the episodic/semantic distinction being dichotomous, it seems quite likely that autobiographical memories instead vary along a 2D continuum depending on the degree to which the memory reflects unique episodic information and/or generic semantic information. Indeed, it is readily apparent in Figure 5 that a great many of our participants' recall responses involved some amount of both reliving and knowledge. As an interesting and relevant alternative to the construct of episodic memories, Rubin and Umanath (2015) proposed the theoretical construct of *event memory* as "the mental construction of a scene, real or imagined, for the past or the future" (see also De Brigard et al., 2022, for a rethinking of the episodic vs. semantic distinction).

Dual-process theory of repetition

Several researchers (e.g., Brewer, 1986, p. 45, 1988, p. 76; Brewer & Pani, 1983, p. 20; Linton, 1982, p. 79; Thompson et al., 1996, pp. 5–9) have hypothesised that repeated autobiographical events tend to produce a generic semantic memory representation at the expense of individual recollective memories (cf. laboratory studies by Bower, 1974; Smith & Handy, 2014).[10] For example, you may retain a schematic generic memory about your daily morning walks to the Starbucks coffee shop with your partner, but not be able to recall the specific conversation or the quality of your hazelnut latte from any particular morning. The correlational data in Table 2 allow a test of the dual-process theory of repetition. If the theory is correct about semantic memory deriving from repeated events, then there should be a positive correlation between the participants' estimates of the *frequency* of occurrence of events in their lives and their *knowledge* ratings. The .36 correlation shown in the table supports the theory. If the theory is correct about recollective memory deriving from single events, then there should be a negative correlation between event *frequency* and ratings of *reliving* (recollection). The −.31 correlation shown in the table supports the theory. Finally, if the theory is correct about semantic memories being constructed *at the expense* of recollective memories, then there should be a negative correlation between ratings of *reliving* and ratings of *knowledge*. The −.55 correlation shown in the table supports the theory. Thus overall, the predicted pattern of positive and negative correlations provide strong support for the dual-process theory of the development of episodic and semantic autobiographical memories.

Metamemory

In addition to the contents, completeness, and accuracy of autobiographical memory for everyday experiences, our study also allowed us to investigate *metamemory* in this context.

Confidence and accuracy. How accurate were participants' metamemory for their everyday experiences? Here we will consider several approaches to analysing the relationship between confidence and accuracy (cf. Roediger et al., 2012), all of which were conducted for the potentially visible response categories, collapsing across retention intervals. As shown in Table 2, there was a medium sized (Cohen, 1992) positive relationship between participants' ratings of confidence during initial recall and their ratings of accuracy during self-scoring, calculated as a repeated measures correlation (Bakdash & Marusich, 2017), $r_{rm}(1075) = .26$, 95% CI [.20, .31], $p < .001$. In the metamemory literature, a commonly used measure of the relative accuracy (or resolution) of metacognitive judgments is the gamma correlation (Nelson, 1984). The overall mean of participant gamma correlations between confidence ratings and accuracy ratings was .43, $SD = .33$, 95% CI [.16, .70], $t(8) = 3.70$, $p = .006$. The individual gamma correlation coefficient was positive for eight of the nine participants, and statistically significant on its own for seven of the nine participants (all positive). The individual coefficients were: −.37, .28, .34, .40, .45, .60, .63, .77, and .77.

The rating data in Table 5 also inform the confidence-accuracy issue. The participants' confidence in their errors (5.4) was a scale unit below their confidence in their correct responses (6.3). A Wilcoxon matched-pairs signed-rank test showed this difference to be significant, $z = 2.31$, $p = .020$. This suggests that participants had some form of metamemory information about the quality of the erroneous response that caused them to reduce their confidence ratings. Overall our data show that participants made good metamemory judgments in the context of everyday autobiographical memory, generally showing higher confidence for more accurate responses.

Metamemory theory of memory confidence. Brewer and Sampaio (2006, 2012) proposed that confidence is an inference based on the products (e.g., a mental image) and processes (e.g., use of recall) involved in a just-completed memory task, along with a set of metamemory beliefs in long term memory (e.g., the belief that recollective imagery is positively related to memory accuracy). Our data speak to this theory.

Processes and products: errors. The data in Table 5 on the rated characteristics of different classes of memory errors show some suggestive patterns about the processes and products involved in the participants' recalls.

The *false assertion of absence* errors certainly stand out. Participants' confidence in these responses was remarkably high, numerically higher than their confidence in the truly accurate responses. Note also the high knowledge and reliving ratings. We suspect that participants were so confident in their recollection of the absence of something they would have otherwise expected, based on their *schemas* for their everyday life, that they went out of their way to mention the absence (cf. Brewer & Treyens, 1981, p. 224).

The *time shift* errors showed confidence nearly as high as the accurate recalls. In these cases, participants recalled something that really did occur, just at a different time. Thus the same processes were likely at play that lead to confidence in the accurate recalls. However, this does not explain the reliving ratings being somewhat lower.

The *intrusion* errors showed the lowest confidence and reliving ratings, along with high knowledge and frequency ratings. This pattern suggests the errors were schema-based (semantic; cf. Barclay, 1986), and that participants had some suspicion of the process.

Metamemory beliefs. What were participants' *beliefs* about the relationship between memory processes and accuracy? Based on the correlations with confidence in Table 2, participants appeared to believe that the experience of reliving indicates accuracy (r_{rm}(1075) = .35, 95% CI [.29, .40], $p < .001$), and that recalls based on more general knowledge are dubious or even misleading (r_{rm}(1075) = −.07, 95% CI [−.13, −.01], $p = .029$). The reliving belief is consistent with findings of positive correlations between recollective experience and confidence across a wide variety of tasks (e.g., sentence memory: Brewer & Sampaio, 2006; event memory: Morris, 1992; Robinson et al., 2000). The relationship between knowledge and confidence has received only a little attention (cf. Sampaio & Brewer, 2009; Rubin et al., 2003) and deserves further investigation.

Validity of metamemory beliefs. How valid were these metamemory beliefs? As shown in Table 2, reliving was only just slightly, if at all, correlated with accuracy (r_{rm}(1075) = .05, 95% CI [−.01, .11], $p = .076$), and knowledge was decidedly not correlated with accuracy (r_{rm}(1075) = .01, 95% CI [−.05, .07], $p = .679$). Thus, participants' beliefs regarding reliving were overly optimistic but in the right direction, and beliefs regarding knowledge were perhaps slightly too pessimistic but not too far off. However, both of these correlations may be watered down due to each other. Both episodic and semantic memory are at play. The correlation between reliving and accuracy may be low because it was possible for participants to produce accurate responses from their knowledge of their lives in the absence of recollection. And vice versa for the correlation between knowledge and accuracy. Furthermore, the occurrence of high-reliving time shift errors would also weaken a reliving-accuracy correlation.

Overall, the metacognitive data show a moderate positive relationship between confidence and accuracy in everyday autobiographical memory. However, the results shed only a little light on the products and processes used to generate confidence judgments. Participants' apparent beliefs about what yields accurate memory showed only little validity, so the results do not provide a good account of how they generated their reasonably appropriate confidence judgments. More targeted research on the metamemory of autobiographical memory is needed.

Table 6. Summary of key contributions and results.

Section	Key Points
Method	
	• Wearable cameras unobtrusively record unbiased records of everyday experience.
	• Pictures/video can be used to cue memory, and to score initial recall (Figure 2).
	• Self-scoring by participants is viable.
	• A full day's recall can be segmented into timeslices, although there is a balance to cuing a specific time period without giving away too much.
	• Recall can be guided by category cues to aspects of experience (Table 1).
	• There are at least three reasons why participants may give new responses upon seeing their pictures (neglected, reminded, forgotten, Table 3).
Results (after 1-4 week retention interval)	
	• Recall contents reflect the structure of participants' lives (Figure 3).
	• Accuracy: $M = 89\%$ ($SD = 6\%$), supporting partially reconstructive view (Brewer, 1986)
	• There were several distinct classes of errors (Tables 4 and 5), their characteristics suggesting underlying processes (retrieval errors, interference, schema inferences, reconstructive imagery).
	• Completeness: $M = 79\%$ ($SD = 11\%$)
	• Considered with previous studies, the more random the sampling of experiences, the lower the completeness of memory (Figure 4).
	• Pictures reminded participants of everyday experience ($M = 23\%$ increase, $SD = 14\%$), including nonvisual aspects.
	• Responses showed distinction between episodic and semantic autobiographical memory (Figure 5), both of which can yield accuracy. Supports dual-process theory of repetition.
	• Metamemory: modest correlation between confidence and accuracy (r_{rm}(1075) = .26)

General discussion

In this section, we will first discuss methodological contributions, and then recap the highlights of our results. Table 6 provides a summary.

Methodology

In carrying out this study we made a number of methodological advances. First, we showed that it is possible to use wearable cameras to unobtrusively gather naturalistic entire-day records of experience for use in testing everyday autobiographical memory, yielding data on fundamental issues in this field of study. Second, we successfully used timeslice cueing to guide participants to a particular part of their day without giving away too much information, in a large majority of cases. Third, we developed a custom computer interface (Figures 1 and 2) that encouraged participants to recall and rate as much as possible about their experiences as guided by category cues. Fourth, we enabled participants to use their pictures to self-score the accuracy of their recalls and report omissions (with the distinction between neglected, reminded, and forgotten proving informative). Finally, we took a number of precautions to avoid potential ethical issues arising from having participants wear cameras while going about their daily activities.

The use of a wearable camera gave us access to detailed and continuous objective records of participants' everyday experiences, which can be used for both cueing and verification of participants' recall. Thus, memory can be tested without the sampling bias involved in selecting particular unique events to record (as in daily diary studies), and without having to actively involve participants in recording (as in Brewer's, 1988 beeper experience sampling study). The more we can externally and unobtrusively record experience, the better we can evaluate memory, and technology will surely continue to enable more recording options (e.g., audio, location, physiological measures; cf. Mehl & Conner, 2012; Mehl & Pennebaker, 2003; Paxton et al., 2015).

One limitation of the current study was that, although participants were not actively involved in recording, the mere fact of wearing a SenseCam likely altered their everyday experiences somewhat. For example, participants had to be mindful of when to disable the camera, and they often interacted with people who were curious about it. Thus, the two acquisition days were to some extent distinctive from other days, potentially making them more memorable. This did provide a convenient way to cue a particular day in the past (e.g., "The second day you wore SenseCam"), but it also potentially reduced the representativeness of original experiences and subsequent memory performance. Routinely wearing a SenseCam for longer periods would ameliorate the latter concern (Doherty et al., 2009; Gurrin, 2009), but would also sacrifice the cueing advantage. Indeed, one issue that the current study highlights is the subtle challenge of cueing a specific period of everyday experience without revealing too much of what is to be remembered. Now that passive recording technology is enabling the evaluation of the accuracy and completeness of autobiographical memory, the question of how best to cue such memory will likely become the next methodological challenge. The effectiveness of any cueing method may depend on the nature of a given participant's original experiences (e.g., homogeneity).

Comparison of planned record procedures in the study of autobiographical memory. In the Introduction to this paper and the Supplemental Materials, we outlined the use of unplanned records versus planned records in the study of autobiographical memory. We reviewed the planned record procedures used in previous studies and argued that *experience sampling* was superior to daily diaries. Having now explored a third planned record procedure, *wearable cameras*, we will outline the intrinsic advantages and disadvantages of this procedure as compared with experience sampling (e.g., Brewer, 1988).

Wearable camera procedure.

Advantages It provides a rich detailed visual record of the participants' experience from an approximately first-person point of view. The camera operates automatically and the participants are not aware of when a picture is taken, so the record is made with minimal interference with their natural experiences. The record covers a large portion of the events in the lives of the participants on a given day so it is possible to sample for memory testing either at random or in a variety of theoretically motivated ways. The pictures themselves can be used both to cue memory and to score recall.

Disadvantages Nonvisual aspects of experience are not recorded. There is no explicit information obtained about the participants' internal mental states during the original experience and data about these mental states can provide important information for the study of autobiographical memory (see Larsen, 1998 for a thoughtful discussion of the phenomenal qualities of an original event vs. the recollective memory of that event). Visual records may still be ambiguous (e.g., the identity of a person in a picture). Battery and file storage space are limitations; this is especially true of newer cameras, such as GoPro, that record full video and audio.

Experience sampling procedure.

Advantages It enables gathering information about the participants' internal mental states (e.g., thoughts, emotions) during the original experience. It also allows the researcher to gather rating data on the characteristics of the original experiences as they happen (e.g., frequency and significance of an event). The narrative format the participants use in describing the original experience may contain useful information (e.g., if a participant states that someone "was not present" then that is good evidence that one would normally expect that person to be present, see Brewer & Treyens, 1981, pp. 222–226).

Disadvantages The random signal (e.g., beeper) draws special attention to momentary experience, and any ongoing activities are interrupted by the requirement to record those activities in narrative form. Even though the instructions tell participants not to treat the event in any special way (e.g., rehearse it), it seems likely that participants may engage in some intentional learning activities. All of these actions have the potential to modify the original memory representation. That said, the data from Thompson (1982) suggest that over the long time intervals typically used in these studies, any such effects do not appear to be very strong.

Assessment. Given the different strengths and weaknesses of the wearable camera and experience sampling procedures, we think developing some form of hybrid procedure would be an important goal for future researchers.

Overview of results

Now we will summarise the findings generated by our new methodology. Figure 6 shows a visual summary of the data. The pie charts show, for the potentially visible response categories, the mean percentages of various types of responses made in both the initial recall phases and the self-scoring phases, collapsed across testing sessions and averaged across participants. The large pie on the left represents everything that a participant could tell us about a particular timeslice on average, whether accurate or not, and both before and after seeing the pictures. The upper-right pie represents omissions participants reported after seeing their pictures. They classified these omissions as: things they had remembered before seeing the pictures but had just *neglected* to mention, things they were *reminded* of, and things they still did not remember (*forgotten*). The lower-right pie represents our classifications of participants' errors. Thus Figure 6 gives a visual overview of the proportion of responses that were accurate, omitted (neglected, reminded, forgotten), and errors (of various types). We will now briefly recap each of the major subsections of our results.

Contents of autobiographical memory. As shown in Figure 3, the most frequently recalled contents of everyday experience were: locations, actions, clothes & belongings, people, visual & spatial, and environment/weather. These results are consistent with Brewer's (1988) findings using a beeper-based experience sampling methodology. He argued that such findings reflect not so much memory phenomena, as the structure of the lives of typical American undergraduates (pp. 75, 77). As a thought exercise to support this view, imagine that one of our participants had decided to become a hermit for a semester. The contents of their recall would include few if any responses in the "people" category.

Omitted responses. By giving participants the option of reporting to us their omitted responses after seeing their pictures, and classifying them (neglected, reminded, forgotten), we showed that these three types of omissions appear to represent three different psychological processes (Table 3, Supplemental Materials).

Nonvisual remindings. Our data show that viewing photos taken by one's wearable camera can prompt high-confidence remindings of even nonvisible aspects of original experience. This result is consistent with Brewer's (1996) proposal that a *recollective memory* often includes aspects of all the original ongoing mental processes including nonvisible mental phenomena such as thoughts and feelings.

Completeness of autographical memory. How much of everyday experience do we remember? This is a question so basic, yet so complicated to answer that few attempts have been made. Our data allowed us to contribute an estimated upper bound on completeness of

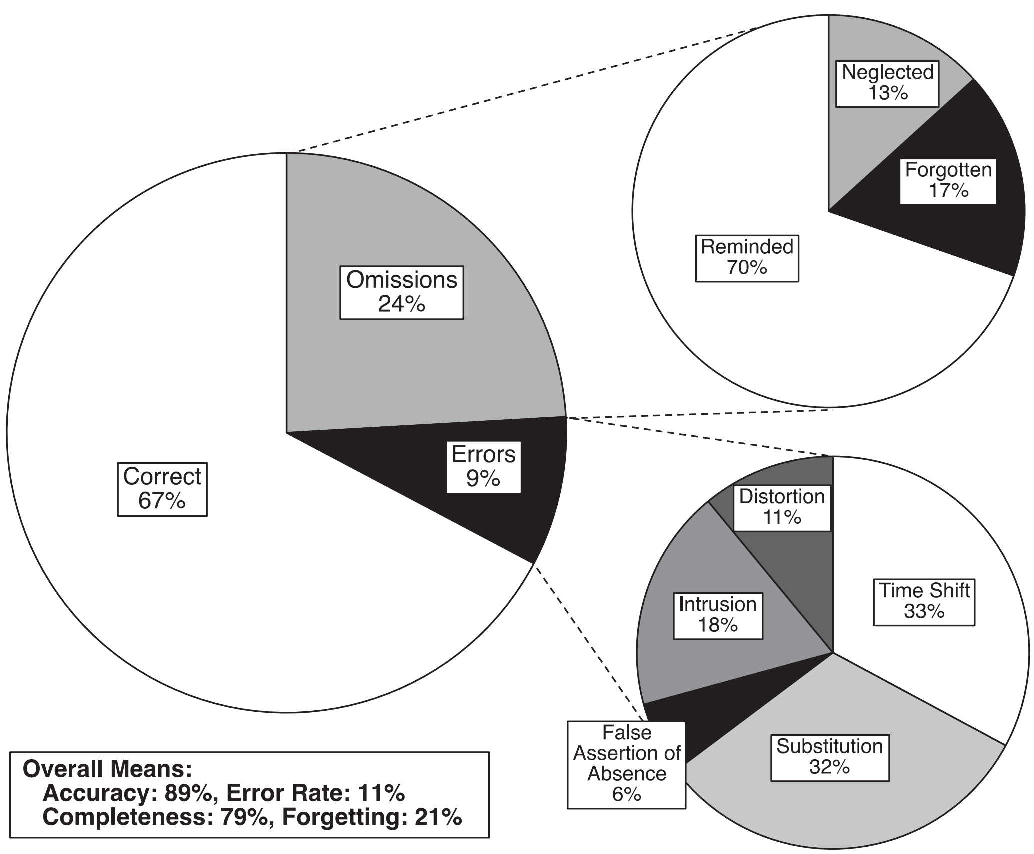

Figure 6. Overview of results.

Note: Mean percentages of various types of responses, collapsed across testing sessions and averaged across participants, for the potentially visible response categories only. Correct and error responses were reported during initial recall. Omissions were reported after viewing pictures. Excluded from this figure: miscued timeslices (see Data Preparation), omissions classified as "other" (see Three Different Types of Omissions), and incorrect initial recalls that could not be classified as errors (see Accuracy of Autobiographical Memory). Accuracy is: correct/(correct + errors). Completeness is: (correct + neglected omissions)/(correct + all omissions).

autobiographical memory for everyday experiences, based on how much participants could report before and after seeing their pictures: an average of 79% completeness, after one week to one month. Stated differently, an average of at least 21% forgetting. Future studies will provide more data to proceed from this starting point.

Retention curves in autobiographical memory. Although our study was not specifically focused on the retention of autobiographical memory over time, we were curious to see how our data fit into the larger picture of other studies with data on this issue. Figure 4 shows that in studies in which the participant selected distinctive or memorable events, the data show remarkably little memory loss over very long time periods (years). However, in studies that selected events at random, the data show some memory loss over weeks and months (more consistent with the typical shape of a forgetting curve), and overall retention that is markedly lower than the distinctive event studies yet nevertheless impressively high (>50%).

Brewer (1988) compared the self-chosen most "memorable" event of the day to randomly sampled events and found that the memorable events had a low frequency of occurrence in the lives of the undergraduate participants. Brewer (1988, pp. 75–76, 1992, pp. 297–298) suggested that lower frequency events tend to yield a more *distinctive* memory representation, which leads to higher retention. Our data support this idea in that the frequency ratings of accurately recalled details were lower than those of forgotten omissions (Table 3).

Memory increase due to pictures. How much more do pictures help us remember of everyday experience? In a previous study using SenseCam wearable cameras (Finley et al., 2011), we found that memory performance at retention intervals ranging from 1 to 8 weeks was improved when participants had reviewed their pictures at the end of an acquisition day, as compared to doing no such review. Effect sizes for that improvement were $d = 0.38$ for recognition ratings (1–7 scale), and $d = 0.25$ for cued recall response lengths. In the current study, we found that the total amount recalled at retention intervals ranging from 1–4 weeks was increased upon reviewing one's pictures immediately after recall based on verbal timeslice cues alone, $M = 23\%$, $d = 0.30$. Although the methods of our two studies are not directly comparable, we find it encouraging that there are similar effect sizes for the reminding power of picture review, either at the end of an acquisition day or at the time of testing. These results, gathered from healthy participants with unimpaired memory, are also consistent with prior research showing that viewing SenseCam pictures has helped memory-impaired individuals to better recall everyday experiences (Berry et al., 2007; Doherty et al., 2013; Loveday & Conway, 2011a; Silva et al., 2018).

Accuracy of autobiographical memory. Our data on memory errors provide an important contribution to the study of autobiographical memory. The use of a wearable camera gives us objective evidence of *randomly sampled* experiences in the lives of our participants. Our use of the participants themselves to score the accuracy of their initial recall means that the errors have been selected by those with the most relevant knowledge. These data, obtained from the everyday lives of healthy undergraduates, provide an upper bound baseline for researchers concerned about memory accuracy in more difficult areas such as flashbulb memory and eyewitness testimony. One area of further research would be using similar methods with healthy older adults, to provide baselines for older adults experiencing memory impairment. Although studies using laboratory verbal memory tasks show a steady decline in memory with normal aging (Park et al., 2002; Schaie, 2005), we think a challenge in studying *everyday* memory performance across the lifespan is that not just the brain is changing, but also the structure of life (see also Draaisma, 2004).

For our healthy young participants, the overall error rate for scorable responses in the visible categories was 11%. In our view, this summary measure rules out both extreme copy theories and extreme reconstructive theories, while fitting with Brewer's (1986) partially reconstructive view.

What types of errors were there? We developed a scheme that classified errors into five categories: time shift, substitution, intrusions, distortions, and false assertion of absence (see Table 4, and Appendix for further examples). These categories are not necessarily exhaustive, but they certainly help in developing hypotheses about the underlying psychological processes leading to errors in autobiographical memory. We suggest that many of these errors can be accounted for by the same constructs used in laboratory memory studies (e.g., retrieval errors, interference, schema inferences, reconstructive imagery).

Our conclusions about memory accuracy are consistent with everyday observation. In the course of a day, we make many successful decisions based on our autobiographical memories, and this indicates that a substantial part of them must be accurate. On the other hand, when discussing an event in the past with someone who was also there, it is not uncommon to disagree about the recalled events. Such disagreement provides evidence for at least a moderate rate of errors, which is what we found.

Finally, it is worth noting that our methodology provided a highly structured framework for recall, with verbal and time cues for timeslices. In everyday life, to remember the experiences of a previous day often requires generating one's own cues and structure (e.g., using schematic knowledge, or temporal landmarks, Shum, 1998).

Episodic versus semantic autobiographical memory. Our data on reliving and knowledge ratings show a clustering of responses into two categories – strongly episodic and strongly semantic (Figure 5, Supplemental Materials). This supports Brewer's (1986) distinction between episodic autobiographical memories (experienced as reliving of the original event) and semantic autobiographical memories (knowledge of the self), an extension of Tulving's original distinction between episodic and semantic (1972). Furthermore, we found that both kinds of memories were capable of yielding accurate recall.

A number of researchers (e.g., Brewer, 1986, 1988; Linton, 1982; Thompson et al., 1996) have proposed that repeated autobiographical events are transformed into generic semantic memory representations at the expense of individual recollective memories. Three predictions of this dual-process theory of repetition are borne out by our data (Table 2). Event frequency was negatively correlated with reliving (recollection), and positively correlated with knowledge. Finally, reliving and knowledge were negatively correlated.

Metamemory. The measures we gathered allowed us to explore autobiographical metamemory to some degree. Little data are available for the confidence-accuracy relationship in memory for everyday experiences, due to the general lack of objective records needed to determine the accuracy of recall (see Barclay & Wellman, 1986 for one exception; see Scoboria & Pascal, 2016 for an approach that sidesteps the problem). Most of the literature on the relationship between retrospective confidence judgments and memory accuracy focuses on eyewitness memory, flashbulb memory, or simple laboratory memory tasks. In our study, we found a medium-sized positive correlation between participants' confidence and their self-scored recall accuracy for everyday experiences. We were also able to begin exploring the metamemory beliefs underlying participants' modestly accurate metamemory judgments (e.g., the influence of reliving), but more targeted future research is needed on this issue. Participants' overall timeslice ratings showed that they did have insight into how much the pictures were helping them to remember.

Conclusion

Overall, the results of our novel methodology showed that autobiographical memory for even randomly sampled everyday experiences can be fairly accurate and fairly complete after one week to one month. These findings constrain and inform theories of autobiographical memory. In addition, automatically-captured pictures from everyday experience can revive memory even for non-visual information, demonstrating the utility of wearable cameras for both the study and the stimulation of memory in naturalistic contexts.

Notes

1. We were not sure how many timeslices participants would be able to complete within a 2 hr time limit, so we used a random order to increase the representativeness of the timeslices that participants did complete. A chronological order would be ideal if time for testing was not limited.
2. See also Barker's (1965) discussion of *behavior episodes*.
3. The means reported here were calculated as the mean of participant means, collapsing across testing sessions and excluding miscued timeslices.
4. The response category "Etc." was used only once. Upon examining the response, which described checking out materials from a library, we reclassified it into the Actions category.
5. A fourth option for omissions (*other*) was available but was only used three times, and none of those uses was accompanied by an interpretable explanation. Thus we excluded those three omissions from all analyses.
6. The phrase "Proustian moment" has come to broadly mean a sudden emotional rush of memories unexpectedly triggered by a sensory cue. We note however that in Proust's original passage (1919/2002, pp. 60–65), the cue was specifically the taste and smell of a madeleine cake with tea, and that the surge of childhood memories, although initiated involuntarily by that cue, did not fully manifest itself until repeated voluntary efforts by the rememberer to plumb the depths of his being.
7. We also tried two other formulas for completeness that gave similar results. In one formula, rather than only counting initial recalls that were scored as completely accurate, we instead counted initial recalls as weighted by accuracy (rating of 1 = 0, 2 = .17, 3 = .33, 4 = .5, 5 = .67, 6 = .83, 7 = 1): M = .80, SD = .10, 95% CI [.72, .89], range: .65–.99. In another formula, we counted all initial recalls that were not completely wrong (accuracy > 1): M = .81, SD = .10, 95% CI [.73, .89], range: .66–.99.
8. These 62 inaccurate initial recall responses were excluded from the accuracy and error analyses reported in this section, and from Figure 6, because they did not provide enough information for us to classify them into an error class. However, these 62 responses *were* included in all other analyses in this paper that did not require classifying errors.
9. A participant from a pilot phase in our first SenseCam study (Finley et al., 2011) had this same insight after free recalling an entire day then watching their movie of that day: "I realized that I remembered very very little of my day. I could only recall memories that involved what classes I went to. However, those are not really memories, just things I know that I did because I have a regular routine."
10. See also Neisser's (1981) discussion of "repisodes" (repeated episodes) and "repisodic memory" and Barsalou's (1988) discussion of summarised events.

Acknowledgements

We thank Ellen Brewer, Cristina Sampaio, and Pooja K. Agarwal for their helpful comments on an earlier draft of this paper, and Alan Scoboria and Linda Henkel for helpful comments on a presentation. Complete raw data are not available because participants of this study did not give written consent for their data to be shared publicly. Supplemental Materials are available. Author contributions: JF and WB conceived of and designed the project. JF created the software, recruited participants, ran participants, and conducted quantitative data analysis. JF and WB conducted qualitative data analysis, and drafted the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by a grant from Microsoft Research, Ltd., Cambridge, UK [grant number # 2007-066].

ORCID

Jason R. Finley http://orcid.org/0000-0001-5921-8336

References

Ariel, B., Farrar, W. A., & Sutherland, A. (2015). The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. *Journal of Quantitative Criminology*, *31*(3), 509–535. https://doi.org/10.1007/s10940-014-9236-3

Baddeley, A. (2010, June). *Reflections on the study of autobiographical memory*. Paper presented at the conference on Theoretical Perspectives on Autobiographical Memory, Center on Autobiographical Memory Research, Aarhus, Denmark.

Baddeley, A. D., & Hitch, G. J. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance VI* (pp. 647–667). Lawrence Erlbaum Associates Inc.

Bahrick, H. P., Hall, L. K., & Da Costa, L. A. (2008). Fifty years of memory of college grades: Accuracy and distortions. *Emotion*, *8*(1), 13–22. https://doi.org/10.1037/1528-3542.8.1.13

Bahrick, H. P., & Karis, D. (1982). Long-term ecological memory. In R. C. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 427–465). Academic Press.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456.

Barclay, C. R. (1986). Schematization of autobiographical memory. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 82–99). Cambridge University Press. https://doi.org/10.1017/CBO9780511558313.010

Barclay, C. R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory and Language*, *25*(1), 93–103. https://doi.org/10.1016/0749-596X(86)90023-9

Barker, R. G. (1965). Explorations in ecological psychology. *American Psychologist*, *20*(1), 1–14. https://doi.org/10.1037/h0021697

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser, & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193–243). Cambridge University Press. https://doi.org/10.1017/CBO9780511664014.009

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

Berry, E., Browne, G., Hodges, S., Smyth, G., Srinivasan, J., Williams, L., Butler, A., Rrustemi, A., Wood, K., Kapur, N., Wilson, B., Conway, M. A., Moulin, C.J. A., & Williams, H. L. (2006). Stimulating episodic memory using SenseCam. *Abstracts of the Psychonomic Society*, *11*, 56–57. https://sites.google.com/site/sensecam/episodic.pdf

Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B., & Wood, K. (2007). The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychological Rehabilitation*, *17*(4-5), 582–601. https://doi.org/10.1080/09602010601029780

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Bower, G. H. (1974). Selective facilitation and interference in retention of prose. *Journal of Educational Psychology*, *66*(1), 1–8. https://doi.org/10.1037/h0035802

Brewer, W. F. (1986). What is autobiographical memory? In D. Rubin (Ed.), *Autobiographical memory* (pp. 25–49). Cambridge University Press. https://doi.org/10.1017/CBO9780511558313.006

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 University Press. https://doi.org/10.1017/CBO9780511664014.004

Brewer, W. F. (1992). The theoretical and empirical status of the flashbulb memory hypothesis. In E. Winograd, & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 274–306). Cambridge University Press. https://doi.org/10.1017/CBO9780511664069.015

Brewer, W. F. (1996). What is recollective memory? In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 19–66). Cambridge University Press. https://doi.org/10.1017/CBO9780511527913.002

Brewer, W. F., & Lockhart, D. W. (1995, November). *Phenomenal experience in laboratory recognition memory tasks*. Paper presented at the meeting of the Psychonomic Society, Los Angeles, CA.

Brewer, W. F., & Pani, J. R. (1983). The structure of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 17, pp. 1–38). Academic Press. https://doi.org/10.1016/S0079-7421(08)60095-X

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory*, *14*(5), 540–552. https://doi.org/10.1080/09658210600590302

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, *67*(1), 59–77. https://doi.org/10.1016/j.jml.2012.04.002

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230. https://doi.org/10.1016/0010-0285(81)90008-6

Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science*, *8*(5), 592–600. https://doi.org/10.1177/1948550617703170

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73–99. https://doi.org/10.1016/0010-0277(77)90018-X

Burt, C. D., Kemp, S., & Conway, M. (2001). What happens if you retest autobiographical memory 10 years on? *Memory & Cognition*, *29*(1), 127–136. https://doi.org/10.3758/BF03195747

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

Chow, T. E., & Rissman, J. (2017). Neurocognitive mechanisms of real-world autobiographical memory retrieval: Insights from studies using wearable camera technology. *Annals of the New York Academy of Sciences*, *1396*(1), 202–221. https://doi.org/10.1111/nyas.13353

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155. https://doi.org/10.1037/14805-018

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). The Guilford Press.

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

Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*(11), 2305–2313. https://doi.org/10.1016/j.neuropsychologia.2009.02.003

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self memory system. *Psychological Review*, *107*(2), 261–288. https://doi.org/10.1037/0033-295X.107.2.261

Crovitz, H. F., & Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, *4*(5-B), 517–518. https://doi.org/10.3758/BF03334277

De Brigard, F., Umanath, S., & Irish, M. (2022). Rethinking the distinction between episodic and semantic memory: Insights from the past, present, and future. *Memory & Cognition*, *50*(3), 459–463. https://doi.org/10.3758/s13421-022-01299-x

Dobbins, C., Merabti, M., Fergus, P., & Llewellyn-Jones, D. (2014). Creating human digital memories with the aid of pervasive mobile devices. *Pervasive and Mobile Computing*, *12*, 160–178. https://doi.org/10.1016/j.pmcj.2013.10.009

Doherty, A. R., Hodges, S. E., King, A. C., Smeaton, A. F., Berry, E., Moulin, C. J. A., Lindley, S., Kelly, P., & Foster, C. (2013). Wearable cameras in health: The state of the art and future possibilities. *American Journal of Preventive Medicine*, *44*(3), 320–323. https://doi.org/10.1016/j.amepre.2012.11.008

Doherty, A., Pauly-Takacs, K., Gurrin, C., Moulin, C., & Smeaton, A. F. (2009, October). *Three years of SenseCam images - observations on cued recall*. Invited speech at SenseCam 2009, Symposium at the 39th Annual Meeting of the Society for Neuroscience: Neuroscience 2009, Chicago, USA. http://doras.dcu.ie/14981/1/Aiden_Kata_Cathal_Chris_Alan.pdf.

Draaisma, D. (2004). *Why life speeds up as you get older: How memory shapes our past*. Cambridge University Press.

Fawns, T. (2020). Blended memory: A framework for understanding distributed autobiographical remembering with photography. *Memory Studies*, *13*(6), 901–916. https://doi.org/10.1177/1750698019829891

Finley, J. R., Brewer, W. F., & Benjamin, A. S. (2011). The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam. *Memory*, *19*(7), 796–807. https://doi.org/10.1080/09658211.2010.532807

Finley, J. R., & Naaz, F. (2023). Strategic use of internal and external memory in everyday life: Episodic, semantic, procedural, and prospective purposes. *Memory*, *31*(1), 108–126. https://doi.org/10.1080/09658211.2022.2126858

Finley, J. R., Naaz, F., & Goh, F. W. (2018). *Memory and technology: How we use information in the brain and the world*. Springer. https://doi.org/10.1007/978-3-319-99169-6

Finley, J. R., Roush, R., Roberts, P., Pfister, J., Mohsen, S., Hendricks, A., & Vozenilek, A. (2023, April). *Can wearable cameras stimulate non-visual memory?* [Poster presentation]. Midwestern Psychological Association Annual Meeting, Chicago, Illinois. Retrieved from: https://www.jasonfinley.com/pubs/NVM-WC1_Poster_4-5-23jrfB.pdf.

Foley, M. A. (2020). Effects of photographic reviews on recollections of the personal past: A new perspective on benefits and costs. *Review of General Psychology*, *24*(4), 369–381. https://doi.org/10.1177/1089268020958686

Gurrin, C. (2009, October). *I Remember … Experiences of a long-term, everyday, SenseCam wearer*. Invited speech at SenseCam 2009, Symposium at the 39th Annual Meeting of the Society for Neuroscience: Neuroscience 2009, Chicago, USA. http://www.vicon.com/Events/SenseCam/documents/gurrin_sensecam09.pdf.

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Henkel, L. A., Nash, R. A., & Paton, J. A. (2020). "Say cheese!": How taking and viewing photos can shape memory and cognition. In S. Lane & P. Atchley (Eds.), *Human capacity in the attention economy* (pp. 103–133). American Psychological Association.

Hodges, S., Berry, E., & Wood, K. (2011). SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*, *19*(7), 685–696. https://doi.org/10.1080/09658211.2011.605591

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K. (2006). SenseCam: A retrospective memory aid. In P. Dourish & A. Friday (Eds.), *Ubicomp 2006*, LNCS 4206 (pp. 177–193). Springer Verlag. https://doi.org/10.1007/11853565_11

Huang, I. (1997). Recognition of student names past: A longitudinal study with N = 1. *The Journal of General Psychology*, *124*(1), 35–47. https://doi.org/10.1080/00221309709595506

Hunt, R. R., & Reed, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001

Hyman, I. E., Jr., & Loftus, E. F. (1998). Errors in autobiographical memory. *Clinical Psychology Review*, *18*(8), 933–947. https://doi.org/10.1016/S0272-7358(98)00041-5

James, W. (1890). *The principles of psychology* (Vol. 1). Henry Holt.

Jobe, J. B., White, A. A., Kelley, C. L., Mingay, D. J., Sanchez, M. J., & Loftus, E. F. (1990). Recall strategies and memory for health-care visits. *The Milbank Quarterly*, *68*(2), 171–189. https://doi.org/10.2307/3350095

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*(4), 371–376. https://doi.org/10.1037/0096-3445.117.4.371

Kalnikaite, V., Sellen, A., Whittaker, S., & Kirk, D. (2010, April). Now let me see where I was: Understanding how lifelogs mediate memory. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2045–2054). ACM. https://doi.org/10.1145/1753326.1753638

Kelly, P., Marshall, S. J., Badland, H., Kerr, J., Oliver, M., Doherty, A. R., & Foster, C. (2013). An ethical framework for automated, wearable cameras in health behavior research. *American Journal of Preventive Medicine*, *44*(3), 314–319. https://doi.org/10.1016/j.amepre.2012.11.006

Kihlstrom, J. F., Kim, M., & Dabady, M. (1996, November). *Remembering, knowing, and feeling in episodic recognition*. Paper presented at the meeting of the Psychonomic Society, Chicago, IL.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*(3), 297–315. https://doi.org/10.1037/0096-3445.123.3.297

Larsen, S. F. (1998). What is it like to remember? On phenomenal qualities of memory. In C. P. Thomson, D. J. Herrmann, D. Bruce, J. D. Read, D. G. Payne, & M. P. Toglia (Eds.), *Autobiographical memory: Theoretical and applied perspectives* (pp. 163–190). Lawrence Erlbaum.

Larsen, S. F., & Plunkett, K. (1987). Remembering experienced and reported events. *Applied Cognitive Psychology*, *1*(1), 15–26. https://doi.org/10.1002/acp.2350010104

Laursen, L. (2009). A memorable device. *Science*, *323*(5920), 1422–1423. https://doi.org/10.1126/science.323.5920.1422

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, *17*(4), 677–689. https://doi.org/10.1037/0882-7974.17.4.677

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

Linton, M. (1978). Real world memory after six years: An *in vivo* study of very long term memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 69–76). Academic Press.

Linton, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (1st ed., pp. 77–92). W. H. Freeman.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, *35*(5), 409–420. https://doi.org/10.1037/0003-066X.35.5.409

Loveday, C., & Conway, M. A. (Eds.). (2011a). SenseCam: The future of everyday memory research? [Special issue]. *Memory, 19*(7).

Loveday, C., & Conway, M. A. (2011b). Using SenseCam with an amnesic patient: Accessing inaccessible everyday memories. *Memory*, *19*(7), 697–704. https://doi.org/10.1080/09658211.2011.610803

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 390–395. https://doi.org/10.1177/0963721417691356

Mehl, M. R., & Conner, T. S. (Eds.) (2012). *Handbook of research methods for studying daily life*. Guilford Press.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*(4), 857–870. https://doi.org/10.1037/0022-3514.84.4.857

Morris, P. E. (1992). Cognition and consciousness: Presidential address to the British Psychological Society. *The Psychologist: Bulletin of the British Psychological Society*, *5*(1), 3–8.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, *9*(1), 1–22. https://doi.org/10.1016/0010-0277(81)90011-1

Neisser, U. (1982). Snapshots or benchmarks? In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 43–48). W. H. Freeman.

Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd, U. Neisser (Eds.), *Affect and accuracy in recall: Studies of 'flashbulb' memories* (pp. 9–31). Cambridge University Press. https://doi.org/10.1017/CBO9780511664069.003

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95* (1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nestojko, J. F., Finley, J. R., & Roediger, H. L., III (2013). Extending cognition to external agents. *Psychological Inquiry*, *24*(4), 321–325. https://doi.org/10.1080/1047840X.2013.844056

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037//0882-7974.17.2.299

Paxton, A., Rodriguez, K., & Dale, R. (2015). PsyGlass: Capitalizing on Google Glass for naturalistic data collection. *Behavior Research Methods*, *47*(3), 608–619. https://doi.org/10.3758/s13428-015-0586-z

Penfield, W. (1969). Consciousness, memory, and man's conditioned reflexes. In K. H. Pribram (Ed.), *On the biology of learning* (pp. 129–168). Harcourt, Brace & World.

Proust, M. (2002). *In search of lost time Vol. 2: In the shadow of young girls in flower* (J. Grieve, Trans.). Penguin. (Original work published 1919)

Ribot, T. (1881). *Les Maladies de la Mémoire [Diseases of memory]*. Germer Bailliere.

Roberts, R. O., Bergstralh, E. J., Schmidt, L., & Jacobsen, S. J. (1996). Comparison of self-reported and medical record health care utilization measures. *Journal of Clinical Epidemiology*, *49*(9), 989–995. https://doi.org/10.1016/0895-4356(96)00143-6

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied*, *6*(3), 207–221. https://doi.org/10.1037/1076-898X.6.3.207

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15* (1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & DeSoto, K. A. (2015). The psychology of reconstructive memory. In J. Wright (Ed.), *International encyclopedia of the social and behavioral sciences, 2e* (pp. 50–55). Elsevier.

Roediger, H. L., III, Wixted, J. H., & DeSoto, K. A. (2012). The curious complexity between confidence and accuracy in reports from memory. In L. Nadel & W. Sinnott-Armstrong (Eds.), *Memory and law* (pp. 84–117). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199920754.003.0004

Rubin, D.C. (2002). Autobiographical memory across the lifespan. In P. Graf & N. Ohta (Eds.), *Lifespan development of human memory* (pp. 159–184). MIT Press. https://doi.org/10.1017/CBO9780511558313.018

Rubin, D. C. (2005). Autobiographical memory tasks in cognitive research. In A. Wenzel & D. C. Rubin (Eds.). *Cognitive methods and their application to clinical research* (pp. 219–241). American Psychological Association. https://doi.org/10.1037/10870-014

Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, *31*(6), 887–901. https://doi.org/10.3758/BF03196443

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Sampaio, C., & Brewer, W. F. (2009). The role of unconscious memory errors in judgments of confidence for sentence recognition. *Memory & Cognition*, *37*(2), 158–163. https://doi.org/10.3758/MC.37.2.158

Schacter, D. L. (1995). *Memory distortion: How minds, brains, and societies reconstruct the past*. Harvard University Press.

Schaie, K. W. (2005). What can we learn from longitudinal studies of adult development? *Research in Human Development*, *2*(3), 133–158. https://doi.org/10.1207/s15427617rhd0203_4

Schönpflug, W. (1986). The trade-off between internal and external information storage. *Journal of Memory and Language*, *25*(6), 657–675. https://doi.org/10.1016/0749-596X(86)90042-2

Schönpflug, W., & Esser, K. B. (1995). Memory and its "Graeculi": Metamemory and control in extended memory systems. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 245–255). Erlbaum.

Scoboria, A., & Pascal, L. (2016). Dissociating appraisals of accuracy and recollection in autobiographical remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(7), 1068–1077. https://doi.org/10.1037/xlm0000230

Sellen, A., Fogg, A., Aitken, M., Hodges, S., Rother, C., & 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. https://doi.org/10.1145/1240624.1240636

Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, *124*(3), 423–442. https://doi.org/10.1037/0033-2909.124.3.423

Silva, A. R., Pinho, M. S., Macedo, L., & Moulin, C. J. A. (2018). A critical review of the effects of wearable cameras on memory. *Neuropsychological Rehabilitation*, *28*(1), 117–141. https://doi.org/10.1080/09602011.2015.1128450

Smith, M. E. (1952). Childhood memories compared with those of adult life. *The Pedagogical Seminary and Journal of Genetic Psychology*, *80*(2), 151–182. https://doi.org/10.1080/08856559.1952.10533620

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1582 1593. https://doi.org/10.1037/xlm0000019

Soares, J. S. (2023). Comparing functions of and recollection with recently taken and recently deleted smartphone camera photos. *Applied Cognitive Psychology*. *37*(4), 699–710. Advance online publication. https://doi.org/10.1002/acp.4064

Spence, R., & Witkowski, M. (2013). *Rapid serial visual presentation: Design for cognition*. Springer. https://doi.org/10.1007/978-1-4471-5085-5

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 241–245. https://doi.org/10.1037/0278-7393.15.2.241

St. Jacques, P. L., Conway, M. A., & Cabeza, R. (2011). Gender differences in autobiographical memory for everyday events: Retrieval elicited by SenseCam images versus verbal cues. *Memory*, *19*(7), 723–732. https://doi.org/10.1080/09658211.2010.516266

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24* (3), 236–243. https://doi.org/10.1207/S15324796ABM2403_09

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(3), 299–313. https://doi.org/10.1017/S0140525X07001975

Talarico, J. M. (2022). Replicating autobiographical memory research using social media: A case study. *Memory*, *30*(4), 429–440. https://doi.org/10.1080/09658211.2021.1903040

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455–461. https://doi.org/10.1111/1467-9280.02453

Talarico, J. M., & Rubin, D. C. (2007). Flashbulb memories are special after all; in phenomenology, not accuracy. *Applied Cognitive Psychology*, *21*(5), 557–578. https://doi.org/10.1002/acp.1293

Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, *10*(4), 324–332. https://doi.org/10.3758/BF03202424

Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. L. (1996). *Autobiographical memory: Remembering what and remembering when*. Psychology Press. https://doi.org/10.4324/9780203773741

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human?. In H. S. Terrace, J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 3–56). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195161564.003.0001

van den Hoven, E., Sas, C., & Whittaker, S. (Eds.). (2012). Designing for personal memories: Past, present, and future [Special issue]. *Human-Computer Interaction*, *27*(1–2).

Vannini, P., & Stewart, L. M. (2017). The GoPro gaze. *cultural geographies*, *24*(1), 149–155. https://doi.org/10.1177/1474474016647369

Vemuri, S. (2005). *Personal long-term memory aids* (Doctoral dissertation, Massachusetts Institute of Technology). http://dspace.mit.edu/handle/1721.1/30242.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18*(2), 225–252. https://doi.org/10.1016/0010-0285(86)90013-7

White, R. T. (1982). Memory for personal events. *Human Learning*, *1*(3), 171–183.

White, R. T. (1989). Recall of autobiographical events. *Applied Cognitive Psychology*, *3*(2), 127–135. https://doi.org/10.1002/acp.2350030204

White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology*, *16*(5), 603–612. https://doi.org/10.1002/acp.819

Whitten, W. B., & Leonard, J. M. (1981). Directed search through autobiographical memory. *Memory & Cognition*, *9*(6), 566–579. https://doi.org/10.3758/BF03202351

Williams, M., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science*, *5*(2), 87–119. https://doi.org/10.1207/s15516709cog0502_1

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Woll, S. (2002). *Everyday thinking: Memory, reasoning, and judgment in the real world*. Lawrence Erlbaum.

Woodberry, E., Browne, G., Hodges, S., Watson, P., Kapur, N., & Woodberry, K. (2015). The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory*, *23*(3), 340–349. https://doi.org/10.1080/09658211.2014.886703

Appendix

Additional Examples of Different Classes of Errors

Error class	Response category	Initial response	Participant's correction notes
Time Shift	Actions	*I practiced my song in the car*	*I did this the next week*
	Actions	*Lay on my bed.*	*I didn't do this till later*
	Locations	*gas station*	*must have gotten gas another time or day*
	Visual & Spatial	*I was facing the TV from my bed.*	*This happened later*
Substitution	Actions	*Set up sugar, salt, pepper, creamer on each table*	*I was not the one who set out the sugar and creamer*
	Actions	*Flipped through the channels on the TV, watching things*	*Actually watched the movie Wolfman*
	Actions	*Drank water*	*Drank Milk*
	Clothes & Belongings	*Backpack*	*Had purse, not backpack*
Intrusion	Locations	*Living room and kitchen*	*Never went there*
	Actions	*talked to friend on phone – sitting and walking along conrete ledge*	*was sitting the whole time*
	Information Content	*I was playing online games.*	*I did not play online games.*
	People	*[S]*	*[S] wasn't in lab that late that day*
Distortion	Actions	*… feverishly writing down notes, calculating homework problems, …*	*Okay … . maybe I procrastinated a lot more in the afternoon than I thought.*
	Environment/ Weather	*Getting fairly dark*	*It's a lot lighter outside than I thought*
	Environment/ Weather	*sunny.*	*There are clearly clouds in the sky but it is mostly sunny.*
	Visual & Spatial	*I was sitting facing West*	*Southwest, not West*
False Assertion of Absence	Actions	*… exchanged notes in class with either [R] or [N] I think one of them was absent that day.*	*Actually both [R] and [N] were there.*
	Actions	*We were just listning to music driving.*	*I was also texting with my phone and taking pictures.*
	Visual & Spatial	*Either [N] or [R] sat next to me.*	*[R] was sitting next to me but [N] was also there.*
	Visual & Spatial	*dinner table with nothing on it*	*saw the table, but there were things on the table*

Note: See also Table 4. Names of people are replaced with bracketed initials to preserve privacy.