

Memory in the Palm of Your Hand: Smartphone-based Methods for Measuring Memory in the Wild

Elizabeth V. Goldfarb^{1*}, Lila Davachi^{2*}, and Morgan D. Barense^{3*}

Abstract

■ Our memories for personally experienced events are essential to who we are and what we do, yet methodological constraints have limited our ability to sample these real-world memories in the real world. The pervasive use of smartphones and wearable technology provides new opportunities to capture these events in the moment and in the wild. These new opportunities also come with new considerations for

the design and analysis of memory experiments. Gathering lessons across forays into smartphone-based memory capture in our three laboratories, this Perspective aims to highlight important decision points, consider pros and cons of different approaches, and provide some tips and tricks to help researchers explore the wilds of smartphone-based memory. ■

INTRODUCTION

Our knowledge of human memory largely derives from tightly controlled laboratory-based experiments. A central strength of this approach is that participants are exposed to the same experiences (with some modulation from counterbalancing and randomization), enabling precise control over what is experienced, how it is attended to and learned, and how memory is measured. More recently, there has been increased adoption of “naturalistic” stimuli such as movies or narratives (Lee, Bellana, & Chen, 2020), yet these paradigms still preserve the core principle of uniformity by presenting the same stimulus to every individual. Meeting the challenges of dynamic stimuli, these studies have laid important groundwork for the development of causal and computational methods to study event perception and memory (Racchah, Chen, Gureckis, Poeppel, & Vo, 2024).

At the same time, it has long been recognized that autobiographical memory, or the memory for personally experienced, real-world events, is central to a full and accurate understanding of memory operates in everyday life. Autobiographical memories are rich, extended in time, emotionally meaningful, and embedded within dynamic contexts that evolve across hours, days, and months. Memories for these varied and multifaceted experiences profoundly influence how people see themselves and the choices that they make in the future. Critically, these personally relevant and idiosyncratic events cannot be captured using uniform stimuli.

Historically, the study of autobiographical memories posed substantial methodological challenges. First, the

diverse experiences that comprise such memories—from everyday activities to emotionally and personally meaningful episodes and from time alone at home to vacations with family—cannot be simulated within the controlled setting of the laboratory. Second, in the absence of tools capable of capturing experiences as they unfolded, researchers necessarily relied on retrospective reports and in-depth interviews conducted days, months, or even years after the events of interest. This work has been extraordinarily influential, providing critical insights into the structure, phenomenology, and function of autobiographical memory, as well as its role in identity, emotion, and mental health (Barry, Hallford, & Takano, 2021; Prebble, Addis, & Tippet, 2013; Cabeza & St Jacques, 2007). Nevertheless, the reliance on delayed recall has limited the ability to link real-world experience to moment-by-moment encoding processes and subsequent memory dynamics. A comprehensive understanding of memory for real-world events requires real-world sampling of memory.

Recent technological advances—notably the widespread adoption (invasion?) of smartphones and wearable devices—have fundamentally altered this landscape. These tools make it possible to sample experience, context, and behavior at a much more fine-grained scale at or near the time that these events occur. They allow us to capture experiences in multiple ways, including video recordings, pictures, audio recordings, free text, and responses to in-the-moment questions. In so doing, these advances open the door to studying memory in a manner that is both ecologically valid and quantitatively tractable. Importantly, this shift does not replace traditional laboratory approaches; rather, it extends them, allowing established theories and mechanisms of memory to be tested, refined, and expanded in the contexts for which

¹Yale University, ²Columbia University, ³University of Toronto

*All authors contributed equally.

memory ultimately evolved—real-world experience. It also can lead to discoveries about memory in the real world that scientists may choose to bring back into the laboratory to further understand their precise mechanisms.

Although such approaches are relatively new for cognitive neuroscience, other fields of psychology such as clinical (Nahum-Shani et al., 2018; Burke et al., 2017) and social (Laurenceau, DiGiovanni, & Bolger, 2026; Harari et al., 2016) have pioneered traditions of smartphone-based data collection that we can leverage to address questions related to memory. Embarking on this area uncovers many decision points that may be daunting (and certainly were for us). What are the design choices that matter? How do we get reliable data to address our questions?

The idea for this Perspective arose from a symposium of the same title held at the 2025 Annual Meeting of the Cognitive Neuroscience Society. Our goal is to share lessons learned from our three laboratories in the use of smartphones to capture memory “in the wild.” Although our specific goals differed, we all converged on a desire to measure memory for real events in our participants’ lives. For some studies (E. V. G.), we wanted to measure how everyday emotional experiences are remembered (Pratt et al., 2025) and pinpoint differences in memory for clinically meaningful events like substance use. In other studies (M. D. B.), we developed a new smartphone application—HippoCamera—to serve two purposes: memory support for those with memory difficulties and a research platform for the study of memory for everyday, real-life events (Martin et al., 2022). Built on core principles of memory function, HippoCamera guides participants to capture distinctive, multimodal cues of everyday moments and then reinforces them through brief, structured replays—mirroring the brain’s own consolidation process. In still others (L. D.), we used a daily diary approach across several weeks asking participants about various aspect of their daily experiences using smartphones as a tool to gather these details. This allowed us

to conduct prospective autobiographical memory formation experiments in the real world, examining what kinds and quality of everyday experiences promote vivid autobiographical memories (Welch, Schelkun, Gasser, Lockwood, & Davachi, 2025).

We will first highlight important decision points in the design and analysis stages of smartphone-based memory studies and review advantages and limitations of different choices (Figure 1). Next, we will provide use examples of choices that we made to address specific research questions in our laboratories. Finally, we discuss specific considerations, limitations, and opportunities made possible by these methods. Ultimately, we hope to help you avoid some of the obstacles that we encountered and feel equipped to ask questions that were out of reach before smartphones made it possible to study memory in the wild.

KEY DECISION POINTS

Q1: Who Chooses the Events?

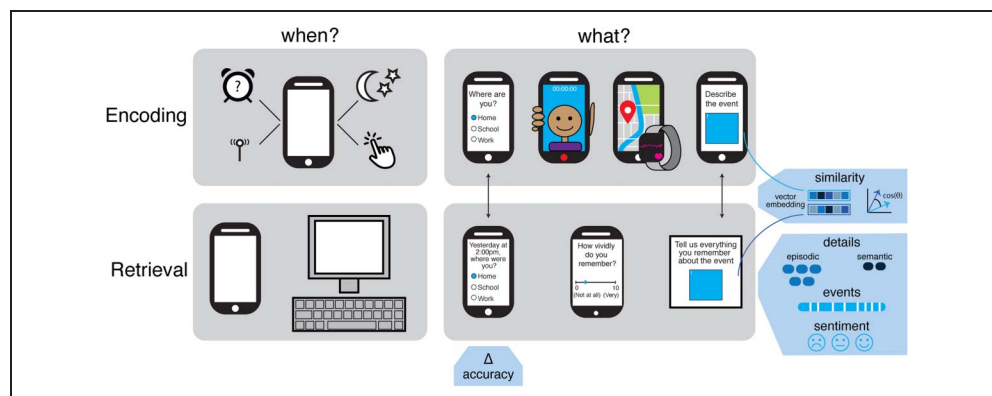
A: The Participant

With this approach, participants decide which experiences they want to report and remember. This can be prospective or retrospective.

In prospective designs, participants are instructed to “record events you want to remember,” with protocols ranging from one to five events per day. One per day was engaging and sustainable, whereas five proved onerous—many days simply do not contain five moments people want to remember, so events became generic and thin on episodic detail. Although this has the benefit of capturing experiences with in-the-moment fidelity, it has the limitation that participants may not anticipate when they are about to embark upon a noteworthy experience.

In retrospective designs, participants reflect at the end of the day on that day’s events. Sharing three events from the current day was acceptable and effective, and participants

Figure 1. Key decision points for smartphone-based memory designs. For encoding, prompts to capture experiences can be delivered at different timepoints. Options include (clockwise from top left) set time windows, end-of-day reflections, initiated by user, or triggered based on sensors. These prompts can also capture different types of information, such as structured questions, video/audio recordings, metadata and location records, and open-ended free text responses. For retrieval, prompts can be delivered within the smartphone environment or in a laboratory setting during a follow-up appointment. The information captured, as at encoding, can include structured questions, self-reflections on memory quality, and open-ended free text responses. The choices of what information is captured, including pairing of responses from encoding to retrieval, can support different types of analysis approaches highlighted in blue.



reported these in a daily diary style data collection. This design has the benefit of catching events that may not have been anticipated as in prospective designs, but has the limitation that the event report is temporally separated, although still accessible, from the experience.

B: The Researcher

In these designs, the researcher prompts the participant to report on the event that is occurring at that moment. This has the benefit of capturing experiences that were not likely to be chosen by the participants, such as everyday events that participants might not consider meaningful enough to select themselves, or aversive experiences that participants may be unwilling to reflect on. Event selection by the researcher could also help avoid intentional and deep encoding or potential memory boosts that can come from a participant deciding, either before or after an experience, that this is something they want to remember and report on. However, this approach is limited as the researcher may send prompts at a time when participants are unable to respond (although this can be mitigated by adjusting how long participants have to reply to a prompt and by keeping the number of questions low) or may miss events of interest by not sending prompts at those moments.

Q2: When is the Event Captured?

A: Random Prompts

This approach is inspired by ecological momentary assessment and experience sampling method techniques. Here, participants receive prompts within a smartphone application asking them to report on events as they are happening. Such prompts are typically set to occur at random times within specified time windows (e.g., morning and afternoon), often customized to an individual participant's wake and sleep times. Random prompts are ideal for designs in which the researcher chooses the events (see Q1, B) and come with the advantages of that approach.

B: User-initiated

In this approach, participants decide, in real time, which moments to record. This induces intentional encoding: People pause, appraise why an event matters, and focus on details—often yielding richer memory. Selection itself also increases daylong attentional monitoring, because participants are actively scanning for moments that merit capture. The same features introduce confounds. Intention-to-remember inflates memory relative to uncaptured events; selection biases toward salient, positive, story-worthy moments; and minimum quotas can generate filler entries. Individual thresholds for “worth remembering” also vary widely. The extent to which intentional encoding presents a challenge for interpretation depends on the research question. For instance, an intervention can occur across some but not all encoded events. Another design

might ask participants to initiate recordings when engaging in different types of activities. In these cases, the comparisons are between different user-initiated events, rather than intentional relative to incidental encoding.

This approach may be appropriate when the goal is to study personally meaningful memories and when any memory boost from the act of capturing is acceptable (or even of interest) or, as above, when comparisons are made between different intentionally encoded events. We also note that this can be combined with A; for instance, some prompts can always be available for participants to initiate, whereas others appear on their phone at random times.

Another approach that allows for experiences to unfold each day as they normally would is to ask participants to share events at the end of the day from that day. This approach draws on daily diary methods from social psychology (Laurenceau et al., 2026; Bolger, Davis, & Rafaeli, 2003). In daily diary paradigms, participants do not interrupt events and experiences in the moment to record or note details. Instead, they reflect back on the day's events each evening and share, diary style, what transpired. An advantage of this approach is that participants can choose events from the day making them more likely to be later remembered. Of course, this blessing is also its downfall in the sense that, because they are already engaging in retrieval of these events, the diary period serves as a re-encoding period as well. There is a sort of trade-off, then, between interrupting or changing an experience in the moment versus retrieving (and possibly changing that experience a little) during the end of day retrieval.

C: Something in Between

Emerging work with sensors (Harari, Muller, Aung, & Rentfrow, 2017) and geotagging suggests that future studies could trigger prompts to capture events that are neither random nor user initiated. Such prompts could occur when a participant is in a particular physiological state or geographic location (Lauckner et al., 2024; note that this is a pretty big hardware ask and can get expensive), or choosing to sample an event based on how participants respond to a particular question (Schneider, Junghaenel, Smyth, Fred Wen, & Stone, 2024).

For participants with memory disorders, automated reminders based on situational analysis (e.g., arrival at a specific salient location) or predesignated events (e.g., set in advance by the researcher or a caregiver) help off-load prospective-memory demands. Sending these alerts at the appropriate time or place reduces missed captures and avoids the need for continuous self-monitoring.

Q3: What Do You Capture When You Record An Event?

A: Multimodal (Video, Audio, Text)

Applications like HippoCamera take advantage of the video and audio recording features of smartphones to

capture rich details. When participants indicate they want to capture an event, HippoCamera prompts them to (i) record a ~24-sec video snippet and (ii) speak a brief ~8-sec verbal description of what happened and what made the event distinctive. The user can optionally add a short title and free-text notes. In the background, HippoCamera combines these inputs into a compact (~8 sec), multimodal cue: a time-compressed video with the spoken description overlaid, linked with any title/notes. Participants are instructed not to worry about producing broadcast- or social-media-ready recordings. The aim is to capture a few distinctive, personally diagnostic details that will trigger recall of the event—that is, a cue to the episode rather than a comprehensive record of it. Think tip of the iceberg, not documentary. Advantages of this approach include multisensory and evocative records of events; limitations include participant discomfort or social awkwardness around recording events in public.

B: Free Text

Participants are asked to write down details of events in free-text responses, with guidance about length and level of detail in these descriptions. Advantages include unique analysis tools available for free-text data (see Q5 for details). In addition to naming and describing events—capturing unprompted, idiosyncratic details that structured questions would miss—they can be asked targeted questions at this time related to specific hypotheses for the study (see C below).

C: Structured Q&A

Experiences can also be captured using structured questions with Likert scales and multiple-choice response options. Limiting reporting to a brief set of structured questions can minimize participant burden, allow for successful event capture in settings when A and B may not be viable, and potentially limit confounds in memory related to deep encoding. This approach can be combined with A and B above to ensure that the researcher has access to key event features (like presence of social others, spatial location, novelty of event, or affective state of the individual) that are hypothesized to be critical for later memory.

D: Sensor Recordings

As mentioned earlier, emerging work with sensors built into smartphones or connected wearables can be integrated into event capture. This would allow for aspects of an experience to be passively recorded, for example, with GPS tracking (Heller et al., 2020; Dennis, Yim, Garrett, Sreekumar, & Stone, 2019).

Q4: How is Memory for These Events Assessed?

A: In-app/Remotely

Many applications allow for “piping” or “looping,” which allow researchers to send content within and between

surveys. This allows researchers to both record events and assess memory for those events within the same platform. For instance, if a participant enters a title for an event at the time of encoding, you can present the same event title in a new survey as a memory cue and then ask them to report what they remember about that event. Applications can also allow the researcher to test participants’ memory for information recorded by sensors, like indicating remembered locations on a map and comparing that to GPS recordings from encoding (McKenzie, Yim, Stone, & Dennis, 2020).

Measuring memory within an app or by sending emails and text messages with survey links can minimize participant burden (as the entire study can be conducted remotely) and can help with standardizing encoding-to-retrieval delays by adding multiple retrieval surveys at specific timepoints relative to encoding. However, concerns about participant burden also limit how deeply you can assess memory for an event and how many questions can be asked.

B: Out-of-app

A major benefit of capturing rich, distinctive multimodal cues is that they let us precisely target specific events later. This enables assessment of routine, everyday experiences—for example, a particular grocery trip or an unexpected encounter—that are hard to identify and probe empirically, even though they make up much of daily life. Because these rich cues provide a clear index of a specific instance, we can revisit that same episode after weeks, months, or years to track how a single autobiographical memory changes over time. More generic cues (e.g., “grocery store trip”) lack the specificity needed for event-level tracking across long delays.

In subsequent sessions—either in the laboratory or over Zoom—memory for these events can be assessed by presenting the memory cue and asking participants to “tell us everything you can remember about the event.” Instructions drawn from the Autobiographical Interview protocol (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002) stress two things: (1) go beyond what appears in the cue and recount the event itself—who was there, what happened, where and when, how it unfolded—and (2) include even seemingly minor details so that someone who was not there could vividly reconstruct the episode. A deceptively simple follow-up “general probe” question (e.g., “Can you tell me any other details about the event?”) is surprisingly helpful in drawing out a few more details.

Q5: What Can You Measure About These Memories?

A: Accuracy

This is an exciting area that is made possible by smartphone recordings of events in the moment. For most autobiographical memories, we cannot determine whether events are remembered accurately—we simply were not there when they occurred. The rare exception is an event

impactful enough to be covered in the media, but such salient events carry their own memory biases, and the news coverage rarely contains enough details to verify a given participant's specific experience. Smartphones change this: we can now know what happened during the event and can measure how much participants' memories deviate from initial reporting.

This is most straightforward with structured questions. In these approaches, participants are asked the same questions at encoding and retrieval. Consistency between encoding and retrieval responses serves as an index of accuracy, with the specific metric varying by question type—binary for categorical responses (e.g., reported location), continuous for sliding-scale responses (e.g., emotional state). These accuracy values for each event feature can also be used to quantify co-accuracy or how much accuracy for one feature depends on another feature (or, put another way, how integrated these different features are in memory; Pratt et al., 2025). They can also be used to exclude unreliable responses (e.g., cases when participants are inaccurate on memory for all event features, which happens rarely but is important to check).

Accuracy can also be quantified by comparing free-text responses between encoding and retrieval. Although techniques like large language models (LLMs) can be applied to any free-text memory data (i.e., even without comparing to an event captured during encoding), new applications are possible with the smartphone-based methods described here. For example, sentence transformer models like Sentence-BERT (Bidirectional Encoder Representations from Transformers; Reimers & Gurevych, 2019) can create vector embeddings of encoded and retrieved descriptions. Then, accuracy can be calculated as the cosine similarity between encoded/retrieved pairs (Welch et al., 2025).

Finally, accuracy can be determined relative to smartphone records such as timestamps and GPS metadata that are captured during an event (Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015). For instance, participants can be prompted to remember where (on a map) and when (on a clock) an event occurred, with the distance between smartphone records and participant reports used to provide measures of accuracy.

B: Quality

Even with photos or video, ground truth is partial. Smartphone media provides anchors—a time-stamped video, a spoken description—but much of what makes a memory rich falls outside the frame: what was said, how it felt, who was just out of view. Consequently, measures of memory quality require interpretation, and we often treat participants' narrative recall as the best available record, recognizing it will carry the embellishments, biases, and inaccuracies characteristic of human memory. With this in mind, detailed descriptions of past events, especially

when cued by rich multisensory recordings, can be analyzed for multiple metrics of memory quality.

Richness, specificity, and episodic content of recall can be ascertained using a scoring protocol adapted from the Autobiographical Interview (Levine et al., 2002). Narrative responses are segmented into internal details (event-specific perceptual, spatial, temporal, and emotional information) and external details (semantic facts, general knowledge, repetitions, or commentary), providing a principled measure of memory quality that goes beyond coarse metrics such as accuracy or RT.

Natural language processing and LLMs are increasingly used to create automated pipelines that extract and quantify features of autobiographical memories from free-text responses. These include event segmentation (identifying where one episode ends and another begins), detail counting (including distinguishing episodic and semantic features), and the emotional sentiment of the memory using analysis tools like VADER (Hutto & Gilbert, 2014; for example, Rouhani, Stanley, COVID-Dynamic Team, & Adolphs, 2023).

In addition to these free-text responses, standardized questions that benefit from postevent reflection (e.g., event significance, event frequency, and location familiarity) can augment understanding of memory quality. Metamemory questions can also probe memory phenomenology, including the vividness, emotional intensity, perspective (own eyes vs. observer), temporal distance, contextualization, and integration with other memories.

METHODS IN ACTION: EXAMPLE USE CASES

How Do People Remember Everyday Emotional Experiences?

Emotional experiences are known to have a privileged place in our memories, yet much of the evidence for this effect comes from extreme personal or newsworthy events or from experiences synthesized to be emotional in the laboratory. How quotidian fluctuations in emotion influence what people remember about everyday experiences remains unclear. To address this question, participants recorded events as they were happening, providing their emotional states in the moment as well as the who, what, when, and where details of these episodes. This design followed an ecological momentary assessment approach, sending participants random prompts within researcher-defined windows and asking them to respond to a brief set of structured questions. These windows were set based on participants' reported sleep and wake-up times. As a memory "tag" that we could use for retrieval (while controlling for variability in the amount of detail participants could include in these tags), participants started each encoding survey by entering the current time (e.g., 12:00 p.m.). The next morning, we prompted participants within the app to tell us what happened (e.g., "Yesterday at 12:00pm..."), using the same questions that we had asked at encoding. This allowed us to limit the

encoding/retrieval delay. For each prompt, participants had a window of 90 min to complete the survey and received reminder notifications after 15, 30, and 60 min if they had not completed it. Within our population of healthy adults (mean age: 24.4 years), participants completed an average of 93% of prompted encoding surveys and 84% of paired memory surveys. They also had above-chance accuracy for individual event features. Using these encoding/retrieval pairs, we calculated the accuracy of memory for each feature and the co-accuracy or integration of memory accuracy between features. We found that events that participants rated as more positive at encoding were remembered with similar accuracy to less emotional events. Notably, these more positive events were also more integrated in memory, indicating that positive emotions facilitated more holistic encoding of these experiences. These findings demonstrate how in-the-moment smartphone sampling can capture meaningful fluctuations in emotional state with consequences for later memory organization (Pratt et al., 2025).

How Does Repeated Reactivation of Everyday Memories Enhance Episodic Richness in Older Adults?

The ability to vividly recollect the recent past declines with age, compromising identity, well-being, and social connection—yet very few interventions specifically target autobiographical episodic memory. To address this gap, older adults used HippoCamera to record and replay everyday events over periods ranging from 2 to 10 weeks, with one to five events recorded per day. A within-subject baseline design was used: Cues were automatically assigned to either a replayed or baseline condition, allowing the benefit of repeated reactivation to be measured while holding encoding constant. Compliance was high across both protocols (Experiment 1: mean = 4.8 cues/day; Experiment 2: mean = 0.95 cues/day). Memory was assessed using an adapted Autobiographical Interview at early and delayed timepoints (up to 3 months postintervention), scored for internal and external details; in-cue details were excluded to ensure scores reflected genuine recollection. Events were matched across conditions on significance, frequency, and memory age. A distinctive feature of this work was that the same cues used to capture everyday events were subsequently used as stimuli during an fMRI scan, allowing us to directly link real-world memory behavior to its neural underpinnings. Participants viewed their own cues and then completed an episodic verification task involving probes that were selected from their Autobiographical Interview narratives. We found that replayed events were recalled with 56–84% more internal details than baseline events, with benefits persisting after a 3-month delay and observed in every participant. Replay also increased hippocampal pattern differentiation, with the degree of differentiation predicting episodic richness on behavioral tests administered outside the scanner. Together, these results demonstrate that user-initiated smartphone cues can serve

as both ecologically valid tools for studying memory in daily life and precisely indexed stimuli for neuroimaging studies of autobiographical memory (Martin et al., 2022).

How Does Experiential Diversity Affect Memory and Well-being during Isolation?

Approximately 24% of older adults experience social isolation, which is associated with increased risk of dementia, depression, and cognitive decline. One underappreciated consequence of isolation is experiential monotony—days filled with repetitive, housebound activities that blur together and are difficult to distinguish as separate episodes. This raises the question of whether engaging in even small deviations from routine could disrupt this monotony and benefit both memory and well-being. To examine this, older adults used HippoCamera to record at least one event per day over 8 weeks during the COVID-19 pandemic lockdowns—a period that provided an unusually uniform experiential baseline against which the effects of individual unique events could be assessed. Participants also completed daily measures of mood, boredom, mindfulness, and perceived passage of time, allowing well-being to be tracked alongside memory. Rather than targeting replay as the manipulation of interest, this study used an event-level uniqueness composite score—combining self-reported event frequency, location frequency, and typicality of day—as the primary predictor of both memory quality and daily well-being. Compliance with daily measures was 89%. Because uniqueness and replay condition were only moderately correlated ($r = .33$), their independent contributions to memory could be disentangled. Unique events were recalled with significantly more internal details, and days with more unique events were associated with greater positive affect, lower boredom, and faster perceived passage of time—effects that were independent of replay. These findings illustrate how the same smartphone platform can be repurposed to investigate questions about experiential diversity, with event-level predictors derived from participant ratings rather than experimental manipulation (Meade et al., 2024).

Which Experiences from Daily Life Are Most Likely to be Vividly Remembered—and Why?

While laboratory studies have long shown that both novelty and prior knowledge benefit memory, they typically treat these as opposing forces, leaving open the question of how their interplay shapes memory for the complex, personally meaningful events of real life. A key challenge is that real-world experiences rarely sit at either extreme: a first visit to a new city and a Tuesday morning commute occupy opposite ends of a novelty spectrum, but most experiences fall somewhere in between. To examine how different types and degrees of novelty influence autobiographical memory, participants completed daily diary surveys for 2 weeks, reporting three events per day and

rating each for regularity (new, periodic, or routine), emotional intensity, and other qualities. Memory was tested approximately 2 weeks later using participants' own event titles as retrieval cues, assessing vividness, episodic detail, and—using sentence-level semantic embeddings—the stability of the memory representation over time. As expected, events labeled as “new” were remembered more vividly and with greater detail than periodic or routine events and showed greater memory stability across time. Critically, however, not all novel events were remembered equally well. Within new events, those that were more semantically similar to the participant's other reported experiences—that is, those with a degree of familiar scaffolding—were recalled with the greatest vividness and were associated with the highest happiness and excitement ratings at encoding. The opposite pattern held for routine events, where greater semantic similarity was associated with lower vividness. Emotional

distinctiveness, computed as the degree to which an event's emotional profile differed from all other events reported by that participant, independently predicted greater memory vividness and episodic detail at recall. Together, these findings challenge the notion that maximal novelty is optimal for memory, instead pointing to a “something old, something new” principle: It is the combination of novelty and semantic familiarity, not novelty alone, that most powerfully supports vivid, lasting autobiographical memory. They also highlight the daily diary paradigm as a powerful method for studying how the texture of everyday experience shapes what we remember (Welch et al., 2025).

Further Questions

In ongoing and recently completed projects, we are asking questions including the following: How does a single night

Table 1. Tips and Tricks

<i>Challenge</i>	<i>Suggestions</i>
How can I label an event so participants know what I am asking them to remember?	– Ask the participant to generate a title for an experience. This will provide a rich memory cue, but might also be too rich if the title itself includes answers to your memory questions. – Alternatives include asking for a “brief” title, giving explicit instructions about what should be included in the title, or by foregoing titles entirely. For example, participants can be asked to type in the current time at encoding, and “Yesterday at XX:XX” can be used as the memory cue. However, this is unlikely to be specific and detailed enough if you are assessing memory after a delay.
How do I match encoding and retrieval data within an app?	– Make sure you know how your survey submission dates are formatted (e.g., MM/DD/YYYY) and that this is coded consistently. – Make sure piped text (e.g., event titles) is saved.
How do I encourage participants to respond to prompts?	– Have a familiar researcher check in regularly with the participant. – Give bonuses, or enter participants in lotteries, based on completion rates. – Minimize burden by using few prompts with many questions or many prompts with few questions (not both!). – Determine when prompts will be sent based on personalized sleep/wake times for each participant.
What software should I use?	Depending on your technical needs, you may be able to leverage an existing application or just email/text participants links to an online survey platform. Smartphone applications designed for ecological momentary assessment have many features that can help support the procedures described here (for a recent list, see Henry et al., 2024). Other guiding questions: – Do you need specialized functionality or direct access to hardware features? – How and where will you securely store data? – What authentication and privacy protections do you need? – How will you ensure compliance with institutional, ethical, and legal requirements (e.g., consent, data retention, regional privacy laws)? – How will you maintain compatibility (e.g., keeping packages, plug-ins, and software dependencies up-to-date and functional with new operating system versions)?

Table 1. (*continued*)

<i>Challenge</i>	<i>Suggestions</i>
What are best practices for building my own software?	– Start small; iterate through rapid development and testing using simple proof-of-concept prototypes. – Build only the core features you need at first. This helps surface technical challenges early and gives you the chance to explore whether those features could be achieved using simpler or alternative approaches. – Leverage established frameworks (e.g., Expo/React Native and Flutter) to gain strong foundations, extensive documentation, and access to developer communities. – Think through supporting infrastructure needs (see above). – If you are working with external developers or a company, or even if development is in-house, ensure you have full access to the codebase, version control repository, and documentation so your team always has the ability of maintaining and updating the application on its own.
How do I extend smartphone-based research to different populations?	– To facilitate equitable inclusion of participants from different socioeconomic backgrounds, provide simple loaner smartphones to participants who do not have access. – Consider your population and anticipate their challenges engaging with the protocol. – Keep the interface simple and intuitive. – Make sure that the manner you are asking them to use the smartphone is not stigmatizing (e.g., calling out a memory deficit). – To account for varied digital literacy, have participants install the application and practice with a researcher in the laboratory or over video call. You could even have a “burn-in” period where data are not analyzed, but used to give participants feedback about the quality of their responses. – Build an FAQ section into your application that participants can access when needed and an easy way to contact the research team.

of sleep shape memory for everyday events? What do people remember about their experiences drinking alcohol? What makes some days—and the experiences within them—more memorable than others? Smartphone studies allow us unique opportunities to ask these questions by capturing many events per participant, recording rich multisensory cues at the time of an event, allowing users to initiate a report when a relevant event is starting, and prompting memory retrieval at specific time windows relative to encoding.

UNIQUE CONSIDERATIONS AND LIMITATIONS

Changing the paradigm for data collection comes with its own set of challenges and considerations. We provide a list of “tips and tricks” in Table 1 and expand on some of these points here.

As with any form of remote data collection, providing clear instructions and maintaining participant engagement can present challenges. Strategies like an extensive onboarding video call to anticipate potential issues, set expectations around event capture, and provide concrete examples of event recording and recall can be very helpful. Financial incentives based on completion rates and regular check-in

phone/video calls with a familiar researcher can also keep participants engaged over long monitoring periods.

In all psychology experiments, there is the potential for selection bias in who participates, and smartphone-based research is no exception. However, smartphone-based data collection also provides unique advantages for inclusion. Remote data collection enabled continued research during the COVID-19 pandemic when laboratory access was impossible—and more broadly, individuals with physical, financial, age-related, or illness-related difficulties that would otherwise preclude in-person participation could use smartphones to have the opportunity to contribute to memory research. This broader participation can provide new insights into individual differences in memory, as well as capturing factors that contribute to fluctuating memory-related processes within an individual (Gillan & Rutledge, 2021). Providing loans of smartphones if needed, together with clear instructions and maintained participant contact as described above, may be a way to level the playing field for participation in memory research.

Finally, shifting away from uniform laboratory-generated stimuli to unique participant-related events raises questions about privacy and confidentiality. We each worked with our institutional review boards to ensure that data acquired from

smartphones are stored securely, with additional protections such as encryption, secure servers, and jittering of GPS coordinates where location data are collected. We also require participants to “opt-in” to initiate recordings and data sharing. With these safeguards, participation involves privacy risks that are comparable to those encountered during typical smartphone use.

CONCLUSIONS

Using smartphones and associated technology to measure memory has been an exciting (dare we say wild?) journey for our laboratories and allowed us to pursue new research questions. Further developments and integration of sensors provide new opportunities to capture events based on participant location and physiological state, as well as to incorporate these dimensions into our understanding of the to-be-remembered event.

There is increasing excitement about using smartphone-based data collection to inform and test models of cognition, including memory-related processes, as well as computations related to mental health based on ecologically valid experiences (Mirea, Nook, & Niv, 2026). We hope this overview helps with thinking about using these methods in your own laboratory as well as the benefits and drawbacks to different design choices.

Acknowledgments

The authors are grateful to Carson Pun at LikeableIT (lead HippoCamera developer) and David Clewett (speaker at our 2025 CNS symposium) for helpful conversations.

Corresponding author: Elizabeth V. Goldfarb, Yale University, New Haven CT, e-mail: elizabeth.goldfarb@yale.edu.

Author Contributions

Elizabeth V. Goldfarb: Conceptualization; Methodology; Visualization; Writing—Original draft. Lila Davachi: Methodology; Writing—Review & editing. Morgan D. Barense: Methodology; Writing—Review & editing.

Funding Information

Centre for Aging + Brain Health Innovation (<https://dx.doi.org/10.13039/501100019829>), grant number: Researcher Clinician Partnership Program Grant. Ontario Research Fund, Research Excellence Award. National Institute on Alcohol Abuse and Alcoholism (<https://dx.doi.org/10.13039/1000000027>), grant numbers: K01AA027832 and R01AA031072. James S. McDonnell Foundation Scholar Award in Understanding Human Cognition (<https://dx.doi.org/10.13039/1000000913>). University of Toronto (<https://dx.doi.org/10.13039/501100003579>), Connaught Innovation Award. Canadian Institutes for Health Research, grant numbers: PJT-126003 and PJT-173336.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article’s gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .333; W/M = .19; M/W = .19; W/W = .286.

REFERENCES

- Barry, T. J., Hallford, D. J., & Takano, K. (2021). Autobiographical memory impairments as a transdiagnostic feature of mental illness: A meta-analytic review of investigations into autobiographical memory specificity and overgenerality among people with psychiatric diagnoses. *Psychological Bulletin*, 147, 1054–1074. <https://doi.org/10.1037/bul0000345>, PubMed: 34968086
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. <https://doi.org/10.1146/annurev.psych.54.101601.145030>, PubMed: 12499517
- Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smailagic, A., et al. (2017). Ecological momentary assessment in behavioral research: Addressing technological and human participant challenges. *Journal of Medical Internet Research*, 19, e77. <https://doi.org/10.2196/jmir.7138>, PubMed: 28298264
- Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, 11, 219–227. <https://doi.org/10.1016/j.tics.2007.02.005>, PubMed: 17382578
- Dennis, S., Yim, H., Garrett, P., Sreekumar, V., & Stone, B. (2019). A system for collecting and analyzing experience-sampling data. *Behavior Research Methods*, 51, 1824–1838. <https://doi.org/10.3758/s13428-019-01260-y>, PubMed: 31240631
- Gillan, C. M., & Rutledge, R. B. (2021). Smartphones and the neuroscience of mental health. *Annual Review of Neuroscience*, 44, 129–151. <https://doi.org/10.1146/annurev-neuro-101220-014053>, PubMed: 33556250
- Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, 11, 838–854. <https://doi.org/10.1177/1745691616650285>, PubMed: 27899727
- Harari, G. M., Muller, S. R., Aung, M. S. H., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, 18, 83–90. <https://doi.org/10.1016/j.cobeha.2017.07.018>
- Heller, A. S., Shi, T. C., Ezie, C. E. C., Reneau, T. R., Baez, L. M., Gibbons, C. J., et al. (2020). Association between real-world experiential diversity and positive affect relates to

- hippocampal-striatal functional connectivity. *Nature Neuroscience*, 23, 800–804. <https://doi.org/10.1038/s41593-020-0636-4>, PubMed: 32424287
- Henry, L. M., Hansen, E., Chimoff, J., Pokstis, K., Kiderman, M., Naim, R., et al. (2024). Selecting an ecological momentary assessment platform: Tutorial for researchers. *Journal of Medical Internet Research*, 26, e51125. <https://doi.org/10.2196/51125>, PubMed: 38175682
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8, 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Lauckner, C., Takenaka, B. P., Sesenu, F., Brown, J. S., Kirklewski, S. J., Nicholson, E., et al. (2024). Combined motivational interviewing and ecological momentary intervention to reduce hazardous alcohol use among sexual minority cisgender men and transgender individuals: Protocol for a randomized controlled trial. *JMIR Research Protocols*, 13, e55166. <https://doi.org/10.2196/55166>, PubMed: 38578673
- Laurenceau, J.-P., DiGiovanni, A. M., & Bolger, N. (2026). Intensive longitudinal methods: Toward a psychological science of daily life. *Annual Review of Psychology*, 77, 513–541. <https://doi.org/10.1146/annurev-psych-040325-025418>, PubMed: 41105835
- Lee, H., Bellana, B., & Chen, J. (2020). What can narratives tell us about the neural bases of human memory? *Current Opinion in Behavioral Sciences*, 32, 111–119. <https://doi.org/10.1016/j.cobeha.2020.02.007>
- 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, 677–689. <https://doi.org/10.1037/0882-7974.17.4.677>, PubMed: 12507363
- Martin, C. B., Hong, B., Newsome, R. N., Savel, K., Meade, M. E., Xia, A., et al. (2022). A smartphone intervention that enhances real-world memory and promotes differentiation of hippocampal activity in older adults. *Proceedings of the National Academy of Sciences, U.S.A.*, 119, e2214285119. <https://doi.org/10.1073/pnas.2214285119>, PubMed: 36512503
- McKenzie, A., Yim, H., Stone, B., & Dennis, S. J. (2020). Using experience sampling to investigate affect at encoding and episodic memory. In S. Denison, M. L. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1660–1666). Austin, TX: Cognitive Science Society.
- Meade, M. E., Chang, M., Savel, K., Hong, B., Martin, C. B., & Barense, M. D. (2024). Unique events improve episodic richness, enhance mood, and alter the perception of time during isolation. *Scientific Reports*, 14, 29439. <https://doi.org/10.1038/s41598-024-80591-z>, PubMed: 39604487
- Mirea, D.-M., Nook, E. C., & Niv, Y. (2026). Cognitive modeling of real-world behavior for understanding mental health. *Trends in Cognitive Sciences*, 30, 350–363. <https://doi.org/10.1016/j.tics.2025.07.009>, PubMed: 41006085
- Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., et al. (2018). Just-in-Time Adaptive Interventions (JITIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52, 446–462. <https://doi.org/10.1007/s12160-016-9830-8>, PubMed: 27663578
- Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences, U.S.A.*, 112, 11078–11083. <https://doi.org/10.1073/pnas.1507104112>, PubMed: 26283350
- Pratt, J. G., Wemm, S. E., Harris, B. B., Huang, Y., Sinha, R., & Goldfarb, E. V. (2025). Positive affect amplifies integration within episodic memories in the laboratory and the real world. *Learning & Memory*, 32, a053971. <https://doi.org/10.1101/lm.053971.124>, PubMed: 39870487
- Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, 139, 815–840. <https://doi.org/10.1037/a0030146>, PubMed: 23025923
- Racah, O., Chen, P., Gureckis, T. M., Poeppel, D., & Vo, V. A. (2024). The “Naturalistic Free Recall” dataset: Four stories, hundreds of participants, and high-fidelity transcriptions. *Scientific Data*, 11, 1317. <https://doi.org/10.1038/s41597-024-04082-6>, PubMed: 39627263
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Hong Kong, China: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>
- Rouhani, N., Stanley, D., COVID-Dynamic Team., & Adolphs, R. (2023). Collective events and individual affect shape autobiographical memory. *Proceedings of the National Academy of Sciences, U.S.A.*, 120, e2221919120. <https://doi.org/10.1073/pnas.2221919120>, PubMed: 37432994
- Schneider, S., Junghaenel, D. U., Smyth, J. M., Fred Wen, C. K., & Stone, A. A. (2024). Just-in-time adaptive ecological momentary assessment (JITA-EMA). *Behavior Research Methods*, 56, 765–783. <https://doi.org/10.3758/s13428-023-02083-8>, PubMed: 36840916
- Welch, E., Schelkun, V., Gasser, C., Lockwood, K., & Davachi, L. (2025). Rich autobiographical memory benefits from both novelty and similarity to other daily experiences. *Scientific Reports*, 15, 45237. <https://doi.org/10.1038/s41598-025-29036-9>, PubMed: 41286312