CrossMark

Autobiographical recall of personally familiar names and temporal information in e-mails: An automatic analytic approach using e-mail communications

Abhilasha A. Kumar[1] · Sudheendra Hangal[2] · Allyson C. Rosen[3]

Abstract

An important question that arises from autobiographical memory research is whether the variables that influence memory in the laboratory also drive memory for autobiographical episodes in real life. We explored this question within the context of e-mail communications and investigated the variables that influence recall for personally familiar names and temporal information in e-mails. We designed a Web-based program that analyzed each participant's year-old sent e-mail archive and applied textual analysis algorithms to identify a set of sentences likely to be memorable. These sentences were then used as the stimuli in a cued recall task. Participants saw two sentences from their sent e-mail as a cue and attempted to recall the name of the e-mail recipient. Participants also rated the vividness of recall for the e-mail conversation and estimated the month in which they had written the e-mail. Linear mixed-effect analyses revealed that recipient name recall accuracy decreased with longer retention intervals and increased with greater frequency of contact with the recipient. Also, with longer retention intervals, participants dated e-mails as being more recent than their actual month. This telescoping error was moderately larger for e-mails with greater sentiment. These findings suggest that memory for personally familiar names and temporal information in e-mails closely follows the patterns for autobiographical memory and proper-name recall found in laboratory settings. This study introduces an innovative, Web-based experimental method for studying the cognitive processes related to autobiographical memories using ecologically valid, naturalistic communications.

Keywords Autobiographical memory · Text mining · Natural language processing

People differ considerably in the ways they experience everyday events and, importantly, in the ways they encode and retrieve those events. Several studies assessing real-world autobiographical memory have demonstrated associations between performance on standardized autobiographical memory tasks and individual difference markers such as age (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002), personality

(Rubin & Siegler, 2004), and self-reported memory abilities (Palombo, Williams, Abdi, & Levine, 2013). Other lab-based and diary studies have also revealed important information about the nature of autobiographical memory, such as the autobiographical forgetting curve (Rubin & Wenzel, 1996), distinctions between episodic and semantic autobiographical memory (Levine et al., 2002), and the positive influence of emotional intensity on autobiographical memory (Talarico, LaBar, & Rubin, 2004).

Despite the vast body of literature on autobiographical memory, our understanding of its nature and the variables that influence it remains biased because of two important limitations. First, most studies on autobiographical memory have focused on events recorded by participants in laboratory settings, through structured questionnaires or interviews. Participants are typically asked to actively report personally experienced events from different life periods (Borrini, Dall'Ora, Della Sala, Marinelli, & Spinnler, 1989; Kopelman, Wilson, & Baddeley, 1989; Levine et al., 2002) or different life contexts, such as meeting a new person or a

✉ Abhilasha A. Kumar
abhilasha.kumar@wustl.edu

1 Department of Psychological & Brain Sciences, Washington University, St Louis, MO, USA

2 Department of Computer Science, Ashoka University, Sonipat, Haryana, India

3 Department of Psychiatry, Stanford University, Stanford, CA, USA

🍦 Springer

family trip or gathering (Piolino, Desgranges, Benali, & Eustache, 2002), following which they are tested on either general recall or specific details about the event, such as its location or time. However, the extent to which findings from laboratory-based studies in fact generalize to real-life events, which are arguably less constrained, remains unknown. It is indeed possible that people's memories about specific details of an event may vary as a function of the specific characteristics of the event itself—for example, the people involved (Maddock, Garrett, & Buonocore, 2001; Sugiura et al., 2006) or the salience of the event—and importantly, may depend on which event they choose to report (Wagenaar, 1986).

Second, practical constraints have limited the assessment of long-term autobiographical memory in naturalistic settings. Laboratory studies typically test autobiographical memory in "snippets"—that is, through discrete life events spread across life periods. Diary studies suffer from a selection bias, such that the events chosen to be recorded are those that are more salient, and the act of recording itself may lead to intensive cognitive restructuring (Wagenaar, 1986). More recent digital diaries or life-logs have made considerable progress in recording naturalistic autobiographical memory and have demonstrated that digital visual records, in fact, produce greater recall than events recorded in written diaries (Hodges et al., 2006). In a study of a patient with severe memory impairment, Berry et al. (2007) showed that the memory for events recorded via SenseCam, a wearable camera device that automatically takes pictures, was far greater than memory for either events recorded in a written diary or events that were not recorded at all. However, the act of recording and viewing images captured through digital diaries or life-logs such as SenseCam may introduce demand characteristics and alter the process of recollection. Other work in this area has attempted to address this bias, by using smartphones that passively record visual, auditory, and location-based information at regular intervals (Dennis et al., 2017; Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015), although the participants in these studies later viewed and segmented the images recorded by the device into distinct episodes, likely leading to rehearsal and restructuring of the events.

In the present study we attempted to address these limitations by investigating the nature of autobiographical memory in a naturalistic setting via e-mail communications. E-mails provide an opportunity to study autobiographical memory-related phenomena by using a time-stamped record of prospectively collected, personally relevant communications. Importantly, given that the participants in this study were unaware (at the time of sending the e-mails) that their e-mails would be analyzed and their memory for these e-mails would be tested at a later point, this study provides an ecologically valid, unbiased sample with which to investigate the relationship of memory for e-mail communications and personally relevant variables, free of demand characteristics. We extracted sentence cues from over 58,000 e-mail communications spread across a year across 44 participants, allowing us to measure autobiographical memory for everyday e-mail communications.

We developed a Web-based program that identified a set of potentially memorable sentences from the participants' year-old sent e-mail archive, using automated textual analysis algorithms based on the open-source MUSE program (Hangal, Lam, & Heer, 2011). These sentences were then used as cues in an autobiographical memory task, in which participants attempted to retrieve the name of the e-mail recipient, similar to a cue-word technique (Crovitz & Schiffman, 1974). Additionally, we collected several other additional variables of interest, including trial-level reaction time, sentiments in the e-mail, cue length, e-mail thread characteristics, and so forth, which provided further opportunity for exploratory analysis on memory for e-mails. Although the primary purpose of this study was to explore whether previously identified patterns of autobiographical memory from offline laboratory settings would generalize to naturalistic, online e-mail communications, we also hoped to address three specific questions: (1) whether memory for the name of the e-mail recipient decreases with increased retention interval; (2) whether any other variables, above and beyond retention interval, explain temporal dating errors for e-mail communications; and (3) whether frequency of e-mailing—that is, contact frequency—predicts recall for the e-mail recipient or the magnitude of the dating error.

Previous research has shown that memory for autobiographical events decreases with an increase in the retention interval. For example, in a seminal diary study on autobiographical memory, Wagenaar (1986) recorded 2,400 events from his life, spread across 4 years, and tested his recall for the events before, during, and 1 year after recording the events. Each event was described using four dimensions: who, what, where, and when. Recall of the event was tested by providing one of the dimensions as a cue (e.g., "who," Leonardo Da Vinci) and attempting to produce the remaining dimensions (e.g., "what," visit to see *The Last Supper*; "where," in a church in Milan; and "when," September 10, 1983). Wagenaar demonstrated that his autobiographical memory decreased over time, as had been previously described using a power function (Rubin & Wenzel, 1996) or fragility function (Rubin, 1982). Other research has also found similar patterns for the retention function. For example, Anderson and Schooler (1991) analyzed newspaper articles from the *New York Times*, parental speech from the CHILDES database (MacWhinney & Snow, 1985), and one of the authors' e-mail communications to show that the likelihood of a memory being needed over time satisfies a power function for retention interval and shows a linear relationship with its frequency of occurrence. They also found that retention interval and frequency produced additive effects on the

memory for the event, suggesting that retention interval and frequency have independent influences on memory.

More recently, Piolino, Desgranges, Benali, and Eustache (2002) tested adults between 40 and 79 years of age with an autobiographical questionnaire that asked participants first to provide general autobiographical information, based on four topics: the names of three personal acquaintances, information about scholastic and professional environments, an important date, and their personal address. Then, participants recalled specific events related to four topics: a meeting or event linked to a person or school, a professional event, a trip or journey, and a family event. Recall was tested using a test–retest procedure in which the participant had to reproduce the general or specific event information on the basis of a cue from the original answer. Consistent with Wagenaar (1986), they showed that memory for events deteriorated greatly with longer retention intervals, suggesting that episodic memory systematically declines over time. If memory for e-mail communications behaves like autobiographical memory episodes, we would expect to see a decline in memory for the name of the e-mail recipient over the period of a year in our study. After they had attempted to recall the recipient name, we also asked participants to rate the vividness of the e-mail conversation, since previous research had demonstrated that vividness modulates the differences between neural networks that govern recent and remote autobiographical memories (Sheldon & Levine, 2013). We predicted that recent e-mail conversations would be rated as more vivid. Furthermore, we also explored the impact of sentiment on recipient name accuracy, and predicted that higher sentiment scores (coded by our algorithm) would lead to greater recall for the recipient, on the basis of previous studies that have shown that emotional intensity enhances autobiographical memory (Talarico, LaBar, & Rubin, 2004).

Another important motivation for this study was to explore the impact of contact frequency on the recall of personally familiar proper names. Previous laboratory studies on the recall of proper names have involved the retrieval of famous faces, places, fictional characters, or films (Condret-Santi et al., 2014; Juncos-Rabadán, Facal, Lojo-Seoane, & Pereiro, 2013; Middleton & Schwartz, 2013) based on short descriptions or definitions. However, in an event-related functional magnetic resonance imaging (fMRI) study, Sugiura et al. (2006) suggested that personally familiar names of people and famous names may have different cortical representations in the brain. In their study, participants viewed images of personally familiar, famous, and unfamiliar people in the scanner and performed a familiar–unfamiliar detection task. The researchers found differential cortical activation during name recognition between personally familiar and famous names, suggesting that accessing information about personally familiar people likely involves retrieving rich episodic information, whereas famous names are represented in a largely semantic context. Additionally,

Brédart, Brennen, Delchambre, McNeill, and Burton (2005) showed that participants were faster at naming familiar people than at making a semantic classification judgment, in contrast to the robust finding in the person identification literature that people perform semantic categorization of a face more quickly than naming the face (Kampf, Nachson, & Babkoff, 2002; Sergent, 1986; Young, McWeeny, Hay, & Ellis, 1986). These studies suggest that personally familiar proper names are represented differently from common names and famous proper names. However, the specific variables that influence the recall for personally familiar names remain unknown, because estimates of familiar name frequency have been difficult to obtain because of a lack of systematic data on familiar name use (Brédart et al., 2005; D'Angelo & Humphreys, 2015). We introduced a novel text-based technique to study recall for personally familiar names. Our algorithm automatically extracted recipient names from e-mails and also recorded the total number of e-mail communications between the participant and the e-mail recipient—that is, an estimate of contact frequency. Thus, we were able to systematically investigate the influence of contact frequency on the recall of the recipient name. If personally familiar names are indeed represented in rich episodic contexts, we predicted that contact frequency would enhance these representations, since previous research has shown that the greater frequency of exposure is related to increased recall for a memory episode (Bluck & Li, 2001).

Additionally, early work on the temporal dating of autobiographical events has been particularly informative about the variables that influence the errors people make while identifying when events occurred. In an experimental study on dating errors by Thompson (1982), young adults recorded unique personal events for themselves and their roommates for a period of 14 weeks. Thompson found that accuracy in dating events decreased systematically with an increase in the retention interval, and that events rated as memorable were remembered better than events rated as low in memorability. The expression "time flies" is commonly used in everyday language, and it often seems "like yesterday" that an event occurred. To explain such systematic misperceptions of time, two terms are used to describe these phenomena and relate them to the hypothesized underlying mental representations of autobiographical timelines: telescoping and time expansion. The phenomenon of *telescoping* is called such because it appears that time *shrinks* toward the present, which is temporally analogous to how distance shrinks when objects are viewed through a telescope (Rubin & Baddeley, 1989). The result of this compressed representation of time is that when people make telescoping errors, events are dated as being more recent than they actually were. Conversely, *time-expansion* errors are called

such because it appears that the underlying mental representation of time *expands*, thus leading to more recent events seeming older than they actually are.

Other work in this area has suggested that time boundaries can impact event dating performance—that is, events closer to the start date of the study (old events) show telescoping, whereas events closer to the end of the study (recent events) show time expansion (Betz & Skowronski, 1997). Both of these phenomena have previously been attributed to a statistical artifact called *regression to the mean* (for a detailed discussion, see Rubin & Baddeley, 1989), which occurs because dating errors at the absolute start of the testing period can only move to a more recent date (causing telescoping) and errors at the absolute end of the testing period can only move to an older date (causing expansion), unless the participant is willing to provide temporal estimates outside the testing period. However, if the temporal dating errors were exclusively attributable to regression to the mean, we would not necessarily expect to find a linear function of telescoping over the entire time period, as has been previously found (Thompson, Skowronski, & Lee, 1988). We would instead expect to see exaggeration at the extremes, when participants come close to the ends of the time window, and specifically, to see a systematic increase in telescoping errors and a decrease in time-expansion errors. To explore the nature of temporal dating errors within e-mail communications, after probing participants for the name of the e-mail recipient in response to the sentence cue, we also asked participants to recall the month in which the original e-mail was written, and measured the temporal accuracy, as well as the temporal distance of the guessed month from the actual month of the e-mail—that is, the temporal dating error. We also classified these dating errors as telescoping or time-expansion errors. Consistent with previous research, we predicted that temporal dating errors would increase with longer retention intervals, specifically for telescoping errors made by participants (Thompson et al., 1988). We also examined the impact of sentiments on temporal dating errors, since emotional memories are often more salient (Kensinger & Corkin, 2004); we hypothesized that sentiment score could interact with memory for temporal information and names in e-mails, but we did not make any a-priori predictions about the direction of the effect, since these analyses were exploratory in nature and we were mainly interested in whether our algorithm could adequately detect sentiments in e-mail.

Method

Participants

Forty-four young to middle-aged (M_{age} = 28 years, SD = 10) adults, including students, faculty, and parents of the students,

were recruited from Ashoka University in India and compensated with Amazon gift vouchers worth INR 500 for their participation. To minimize technical complexity, only participants with Gmail e-mail archives were allowed to participate. We also imposed a screening criterion on the participants' e-mail archive, such that participants were required to have sent at least ten e-mails in each month of the past year, in order to be able to generate a sufficient number of test items for the study. To ensure that we would have adequate power to detect effects, we advertised the study on university e-mail portals and used referrals to recruit as many participants as possible.

Materials

Access to original program An offline version of the study is available to download and run locally from the MUSE Memory Study. Additionally, an instruction manual can be accessed through the supplementary materials or at the GitHub page. The manual describes detailed procedures to run the program and also provides suggestions for modifying some aspects of the original study in order to answer additional questions.

General overview This section broadly describes the experimental paradigm; the complete procedure is described in a subsequent section. Participants opened the Web-based test screen and logged into their e-mail, after which they began the experiment. For each experimental trial, participants viewed two sentences from a sent e-mail and attempted to recall the name of the e-mail recipient. To generate these sentence stimuli, we applied text-based analysis using an empirically derived scoring protocol, described below. Participants also rated vividness of recall for the e-mail conversation and estimated the month in which they had written the e-mail. After completing the test phase, participants completed a postexperiment questionnaire in which they evaluated the errors they had made in the test phase.

Data cleaning After the participant had logged into their e-mail, all the sent e-mails from the past year were fetched to an encrypted, privacy-protected, secure server housed in the Department of Computer Science at Ashoka University in India. E-mail attachments, text formatting, images, and other nontext data were excluded in this first step. The algorithm also removed quoted and forwarded sections of the e-mail by identifying fixed templates used in e-mails.

Pilot testing We developed a novel scoring protocol for identifying the test items, based on multiple rounds of pilot testing on the e-mail archives of ten pilot participants, whose data were not included in the final sample of participants. Each pilot participant's archive was first processed and cleaned and then was used to create a contact address book. Next,

arbitrary constants were used to generate an initial scoring protocol (described below), which was then iteratively refined on the basis of the subjective and objective responses of the pilot participants to the postexperiment questionnaire.

Scoring protocol for generating recipient names Using the cleaned e-mail archive, we performed entity resolution in order to match different e-mail addresses belonging to the same person, in order to generate a contact address book for each participant. Generic mailing lists and recipients e-mailed only once were excluded from further processing. The empirically derived scoring algorithm (based on the pilot testing) was then applied to all remaining recipients in the address book.

For each month of the year, the algorithm identified all recipients who had last been contacted in that particular month—for example, for the month of May, all recipients who had not been contacted after May (via group and individual e-mails) were selected. A *recipient score* was assigned to each of these candidate recipients, proportionate to the total number of e-mails (group and recipient-only) sent to the recipient within that month.

Arbitrary constants were used to compute recipient scores during pilot testing, which were then iteratively adjusted on the basis of whether the pilot participants judged their errors as being valid or invalid in the postexperiment questionnaire. The final parameters used in the scoring protocol are described in Table 1.

Feature extraction After identifying the candidate recipients for each month of the year, we programmed a tokenizer that retrieved all sentences from the participant's most recent e-mail to each of the recipients, along with a set of features. The tokenizer excluded all forwarded e-mail, to ensure that participants had indeed written the sentences themselves, and also excluded extremely long (over 200 characters) or short (less than ten characters) sentences, to control for cases in which participants had simply pasted a large body of text—for example, a book passage or a code fragment—or written a very short reply. We also performed sentence-to-sentence matching, to identify and exclude exactly the same sentences that were used across e-mails, such as "Looking forward to meeting you," "Hope you are doing well," and so forth.

Table 1 Scoring algorithm for choosing the test items

Feature Number	Feature	Parameter	Score
Recipient Score			
1	n: number of e-mails exchanged with recipient in that month	exceeds 1	$n*10$
		exceeds 5	$n*15$
		exceeds 10	$n*20$
Sentence Score			
2	Emoticons	:), :(, :O, etc.	+ 5
3	Question mark	?	+ 7
4	Exclamation	!	+ 7
5	Proper names	Identified by named entity recognizer	+ 10
6	Family words	husband, wife, partner, spouse, sister-in-law, brother-in-law, mother-in-law, fiancé, fiancée, aunt, . . . , neighbor, relative, roommate	+ 10
7	Reflective words	absorb, accept, admit, affirm, analyze, appreciate, assume, convinced of, . . . , trust, understand, vision, visualize, wonder	+ 10
8	Travel words	flight, travel, city, town, visit, arrive, arriving, land, landing, reach, reaching, train, . . . , road, bus	+ 10
9	Emotion words: Positive	happy, alive, understanding, playful, calm, . . . , drawn, confident, hopeful, amazing, fantastic, wow	+ 10
10	Emotion words: Negative	angry, depressed, sad, sadly, . . . , wronged, menaced, alienated, wary	+ 5
E-Mail Score			
11	s: sentiments	based on superlative, congratulations, wow, confidential, memories, family, life event, religion, festivals, love, vacations, racy, emergency, grief, anger	$s*10$
12	x: other sentiments	based on Features 2–10	$x*2$
13	T: number of sent e-mails in thread		$5*T$

Final Score = Recipient Score + Sentence Score + E-Mail Score

The features extracted by the tokenizer included metrics that had been indicative of memorable content in earlier work—that is, MUSE (Hangal et al., 2011)—such as affect words (e.g., "sad," "happy"), emoticons (e.g., ☺, ☹), question marks (?), and exclamation points (!). Specifically, MUSE is an open-source program that has widely been used to analyze communication patterns and sentiment use in e-mails. MUSE uses data-mining and machine-learning algorithms to extract meaningful cues from e-mails that are likely to evoke memories. Specifically, an early version of MUSE extracted group names using co-recipiency, named entities using the Stanford NLP toolkit (Finkel, Grenager, & Manning, 2005), and picture cues from e-mail attachments. Additionally, the algorithm also extracted sentiments in e-mail using an English lexicon consisting of 20 categories, comprising various emotions, life events, expletives, and so forth, that might be memorable, similar to other tools used for sentiment analysis, such as LIWC (Tausczik & Pennebaker, 2010) and SentiWordNet (Baccianella, Esuli, & Sebastiani, 2010). These words were then matched with the contents in the e-mail for feature detection. Hangal, Lam, and Heer had also conducted a preliminary study with six participants to rate the usefulness of each type of cue (group, name, picture, and sentiment) in reviving memory of the particular e-mail episode, on a scale of 1 (*not useful at all*) to 5 (*very useful*). Picture cues were rated highest ($M = 4.25$), followed by name ($M = 4.17$), sentiment ($M = 4.17$), and group ($M = 3.83$) cues. Hangal, Lam, and Heer also reported anecdotal evidence that sentiment and name cues were highly evocative and allowed participants to recollect significant events in their lives. Although the sample size for the study was too small to be conclusive about the usefulness of these cues, modified and derived versions of MUSE have since been widely used to detect sentiments in e-mails (Hangal, Chan, Lam, & Heer, 2012; Nagpal, Hangal, Joyee, & Lam, 2012; Schneider, Chan, Edwards, & Hangal, 2017). In the present study, as is described in Table 1, we extracted sentiments corresponding to 15 categories of emotions, measured the occurrence of

emoticons and exclamation points from the specific e-mail prompt, and also calculated the total sentiments in the e-mail document. These occurrences were then summed in order to produce a sentiment score that was used as a predictor in subsequent analyses.

In addition to extracting sentiment features using MUSE, proper names mentioned in the sentences were extracted by a tokenizer using a derivative of MUSE, the "e-mail: Process, Appraise, Discover, and Deliver" (ePADD) named entity recognizer (Schneider et al., 2017), which excluded common Internet abbreviations (e.g., BTW, FYI), the participant's own name, and the recipient's name from the process. Other extracted features included but were not limited to the number of sentences, the length of the sentences, and the number of sent e-mails in a thread. A sample list of extracted features is presented in Table 1; the complete list of extracted features is available from the authors.

Scoring protocol for generating sentence cues The scoring phase produced two scores: the *sentence score* and the *e-mail score*. All sentences in the e-mail that were extracted by the tokenizer were first segmented into consecutive pairs of sentences and then scored on the basis of the frequency count of features they contained. Sentences were penalized for exceeding or not having a certain number of characters, to avoid presenting extremely long or short sentence cues. The e-mail containing the sentence was scored in terms of the number of affect words and the number of sent e-mails in the thread containing the e-mail. A composite *final score* was calculated by aggregating the recipient, sentence, and e-mail scores. For each month, wherever possible, at least one and at most four items with the highest *final score* were chosen as the final items for that month. The final parameters and complete scoring methodology are described in Table 1. Examples of the sentence cues generated by the program (from one co-author's own e-mail archive) and the corresponding scores for some extracted features are displayed in Table 2.

Table 2 Examples of sentence cues generated by the program

Sentence Cue	Recipient Score	Sentence Score	E-Mail Score
I have revised my AAUW research plan based on your comments and also feedback from Dave Jessica and Emily. I'd be very grateful if you could go through it and give me some feedback on the same!	90 (based on number of e-mails exchanged in that month)	63 (based on 1 exclamation, 5 names including recipient, length penalty, 1 sentiment)	11 (5*2 sent e-mails in thread + 1 sentiment in document)
I just realized that you will be here for a talk tomorrow and it would be lovely to meet and talk if you have some time later in the evening. I'm looking forward to your talk and I hope I get to see you!	23 (based on e-mails exchanged in that month)	21.7 (based on 1 name including recipient, 1 exclamation, length penalty, 1 sentiment)	43 (5*6 sent e-mails in thread + 10*1 sentiment + 3 other sentiments in document)
Sorry for the late reply I just got back to Delhi after a wonderful time in Bombay :) I'll e-mail you the tickets the minute I book them should be done in a day or two. Also would it be okay if my mother stayed with me for the two days I'll be in Bangalore?	120 (based on number of e-mails exchanged in that month)	54 (based on 4 names including recipient, 1 emoticon, 1 question mark, 1 family sentiment, length penalty)	18 (5*1 sent e-mail in thread + 10*1 sentiment + 3 other sentiments in document)

Privacy concerns Due to the sensitive nature of the stimuli in this study, we followed very strict procedures to protect the participants' e-mail content. The software made it possible that at no point during the experiment could any member of the research team view or access any of the participants' personal e-mail messages, and the data were directly fetched to a securely housed encrypted server. All participants accessed the study remotely and were specifically instructed to start the experiment when they were in a secure location—for example, their home, personal office, or other private location. After informed consent, the participant logged into their e-mail account, and a copy of their sent e-mail archive was created and screened to evaluate whether the participant had a sufficiently active account to generate memorable test items. An example of a person who would be excluded would be someone who primarily used their work e-mail for correspondence rather than the personal account they submitted. The copy of the e-mail archive was deleted if the participant failed to meet the screening criteria. For participants who passed the screening phase, the copy of the archive was used to generate sentence cues automatically, without human interaction, and after that process was complete, the original archive copy was deleted. After the participant had completed the study, the sentence cues were deleted, so they were not stored in the final data file created from the experiment, to further protect the content in the e-mails.

Procedure

After informed consent, participants clicked on a Web-based link and logged into their Gmail account. Participants waited while their e-mail archive and demographic information were evaluated to determine whether they would qualify for the study. Upon qualifying, participants were provided a link to the experiment. Following a brief explanation about the questions and a demonstration of the task, participants began the experiment. For each experimental trial, participants saw two sentences from their sent e-mail as a cue and attempted to recall the name of the e-mail recipient. To facilitate recall, the letters in each part of the name were displayed using dashes (e.g., For "Humpty Dumpty":_ _ _ _ _ _ . _ _ _ _ _ _), and also in plain text (e.g., two words [six letters, six letters]). Participants responded to the cue by typing the name in a text box and/or by choosing among five options to specify their retrieval state: (1) The name was easy to recall, (2) I got the name after a while, (3) The name is on the tip of my tongue!, (4) I know the person, not the name, and (5) I don't know. After choosing one of the five options, participants were provided the initials of the name as a second hint (e.g., H _ _ _ _ _ D _ _ _ _ _). Participants could then choose to move on or could attempt to retrieve the name again. Next, participants rated the vividness of the e-mail conversation on a 10-point sliding scale with ratings that ranged from 1 (*no memory*) to 10 (*strong*). Participants also estimated the month in which they had written the e-mail by choosing from a drop-down menu with all the months of the past year as options, as well as "I have no idea" (see Fig. 1).

After completing the test phase, participants completed a postexperiment questionnaire. Participants viewed all the items that the algorithm had marked as incorrect and provided responses about their errors by choosing among five error types: (1) I feel like I should have remembered this name;

Fig. 1 Experiment procedure. Participants viewed a two-sentence cue and typed the name of the e-mail recipient in the text box. They also elaborated on their recollection, specified the vividness for the e-mail conversation, and estimated the month in which they had written the e-mail

(2) My answer is essentially correct; (3) I recognized the context, but not the person; (4) I have trouble remembering this name; and (5) It was a vague/generic sentence. After completing the error judgment phase, participants filled out a postexperiment survey about their experience and closed the program window.

Results

Sentence cue distribution

Figure 2 displays the total number of sentence cues across the retention intervals received by each participant. On average, each participant received 19 sentence cues (SD = 5.5). Although there was considerable variation in the number of sentence cues, due to vast differences in the volumes of participants' e-mail archives, as is shown in Fig. 2, 39 (88.6%) of the participants received one to four cues from each month (M = 2.02, SD = 0.34), spread across a wide range of months (average range = 10.37 months). Five participants had fewer than 12 sentence cues for the entire study, because of a lack of valid sentence candidates in each month, but excluding their data did not affect the interpretation of our results, and hence their data were retained in the final sample.

Error responses

Figure 3 displays the mean occurrence of these response types across participants. The average accuracy for recipient name recall was 48%. Note, however, that the percent occurrence of Error Type 2 ("my answer is essentially correct") is relatively high (18%), because the algorithm only marked answers that exactly matched the recipient name in the address book, and did not consider parts of the name or spelling errors as correct responses. Pilot testing had indicated that participants were fairly accurate at evaluating their responses; thus, all responses marked as Error Type 2 were counted as correct responses. Furthermore, we also excluded all trials that the participants had marked as Error Type 5 ("It was a vague/generic sentence") from further analysis (11.5% of the total responses), since pilot testing had indicated that such responses corresponded to sentence cues that were indeed very vague and generic. In compliance with the institutional review board (IRB) protocol for the study, we could not store any of the sentence cues presented to participants, so a direct review of the test items was not possible, in which case the participant responses from the postexperiment questionnaire were assumed to be fair.

Overview of analyses

All analyses were conducted on two dependent variables: the accuracy of recalling the recipient name from the sentence cue, and the magnitude of the dating error—that is, the distance of the participant's estimate of the month in which the e-mail had been written from the actual month of the e-mail. Recipient name accuracy was coded as a binomial variable (0, 1), and temporal dating error was coded as a continuous variable with a range of 0–12 months. The main predictor variables used in these analyses were retention interval, vividness, contact frequency, and type of temporal dating error (telescoping or time expansion). For each test item, *retention interval* indicated the number of months that had passed since the e-mail containing the sentence cue had been written by the participant, *vividness* indicated the participant's vividness rating for the e-mail conversation on a scale of 1 to 10, and *contact frequency* corresponded to the total number of e-mails sent by the participant to the e-mail recipient over the year. *Type of temporal dating error* indicated whether the participant had guessed the month of the e-mail as being more recent (telescoping) or more remote (time expansion) than its actual month. We also examined the effect of sentiment on recipient name accuracy, on the basis of the sentiments coded by the algorithm in the e-mail and sentence clue (see the Materials section). However, sentiment score did not influence recipient name accuracy. Hence, all reported analyses for recipient name accuracy do not include sentiment as a predictor. We analyzed the effects of the predictor variables on recipient name accuracy using generalized linear mixed models (with a logit link), and the effects of predictor variables on the magnitude of dating error with linear mixed-effect models using the lme4 package (Bates & Sarkar, 2005) in the RStudio environment (R version 3.4.2;, R Core Team, 2017). All data and the analysis scripts are available at https://github.com/abhilasha-kumar/Memory-for-emails.

Recipient name accuracy

We started all analyses with a null model that included the binomial dependent variable (recipient name accuracy) and participants and items as random factors; we added predictor variables to the null model incrementally in order to evaluate whether each predictor improved the fit of the model. Model fit was assessed using chi-square tests on the log-likelihood values to compare the models. Figure 4 displays the model fits for the predicted probabilities of recipient name accuracy as a function of retention interval and person-centered contact frequency, overlaid on the raw naming accuracy data. Table 3 displays the model estimates and coefficients for the best-fitting model. We used grand- and person-mean centering to evaluate the effect of contact frequency on recipient name accuracy. The grand-mean-centered estimate of contact frequency
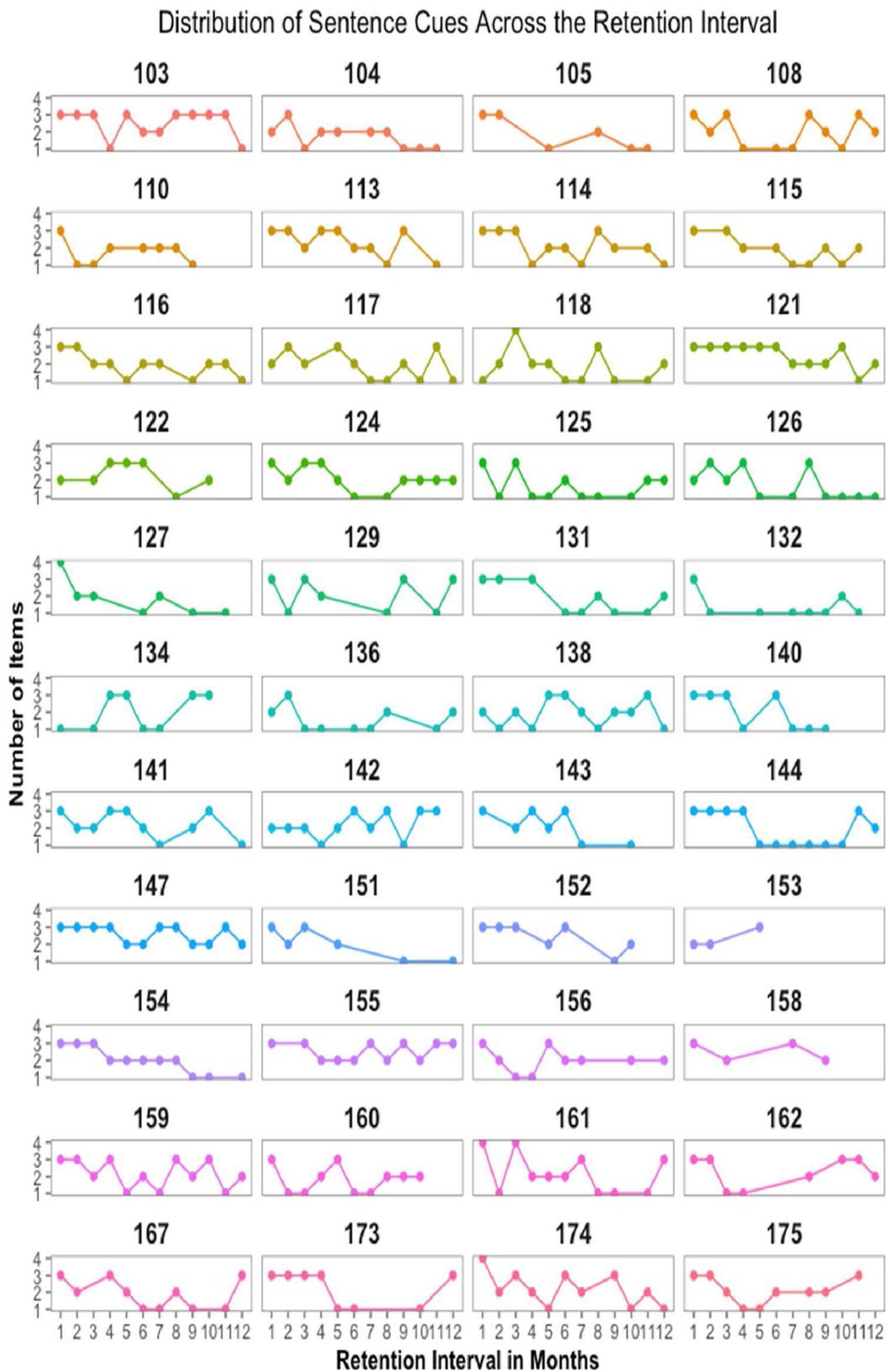
Fig. 2 Distribution of numbers of sentence cues from each retention interval (in months). Each color and number denotes a different participant

corresponded to the average number of e-mails sent by the participant during the year, across all recipients included in the test, and served as an indicator of the participant's general e-mailing frequency. The person-mean-centered estimate of

Mean Occurence of Response Types and Errors
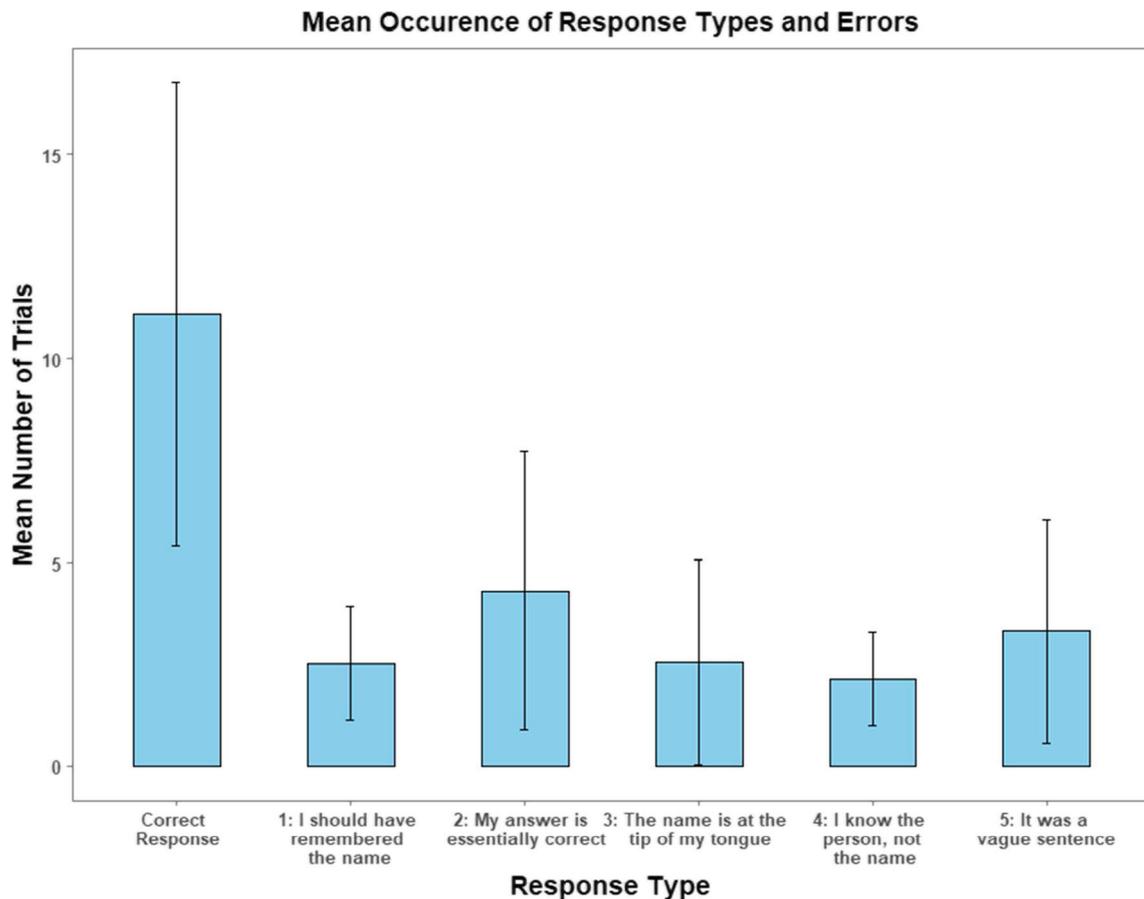
Fig. 3 Average occurrence of each response type across participants. Error bars represent standard errors

contact frequency indicated whether the number of e-mails sent to a particular recipient was above or below the participant's grand mean. Including both types of means as predictors in the model allowed us to answer two different questions: (1) Does general e-mailing frequency predict accuracy for recalling the name of the recipient, and (2) Does e-mailing a recipient above the average e-mailing behavior predict accuracy for recalling the name of the recipient.

Model comparisons indicated that retention interval strongly predicted the likelihood of correctly recalling the name, odds ratio = 0.82, $z = -7.21$, $p < .001$. Interestingly, the average number of e-mails sent by the participant did not predict accuracy, $p = .113$, but the person-centered estimate of contact frequency did predict accuracy, odds ratio = 2.32 [1.43, 4.05], $z = 3.19$, $p = .001$. Furthermore, retention interval and frequency did not interact. We also found that vividness ratings were negatively related to longer retention intervals, $r = -.35$, $p < .001$. These results suggest that participants' memory accuracy for recipient names declined with longer retention intervals, and that e-mails remote in time were also rated as being less vivid. Additionally, the names of recipients who were e-mailed more frequently over the year were recalled more accurately.

Retrieval state declaration

Figure 5 displays the mean numbers of trials on which participants chose each of the five retrieval states across retention intervals, grouped by the month of the e-mail. As is shown in Fig. 5, participants were most likely to choose the option "The name was easy to recall" (in blue) for recent e-mails. The likelihood of choosing this option decreased at longer retention intervals, as was revealed by a correlational analysis of retrieval state (coded as continuous and ranging from 1 to 5) and retention interval, $r = .16$, $p < .001$. There were, however, no differences for the other retrieval states across retention intervals.

Magnitude of temporal dating errors

Before analyzing temporal dating errors (i.e., the distance of participant's estimates from the actual month), we excluded all trials on which the participant did not provide a temporal estimate for the month and all trials for which the temporal estimate provided was correct. We started all analyses with a null model that included the continuous dependent variable (magnitude of temporal dating error) and participants and items as random factors; we added predictor variables to the null model
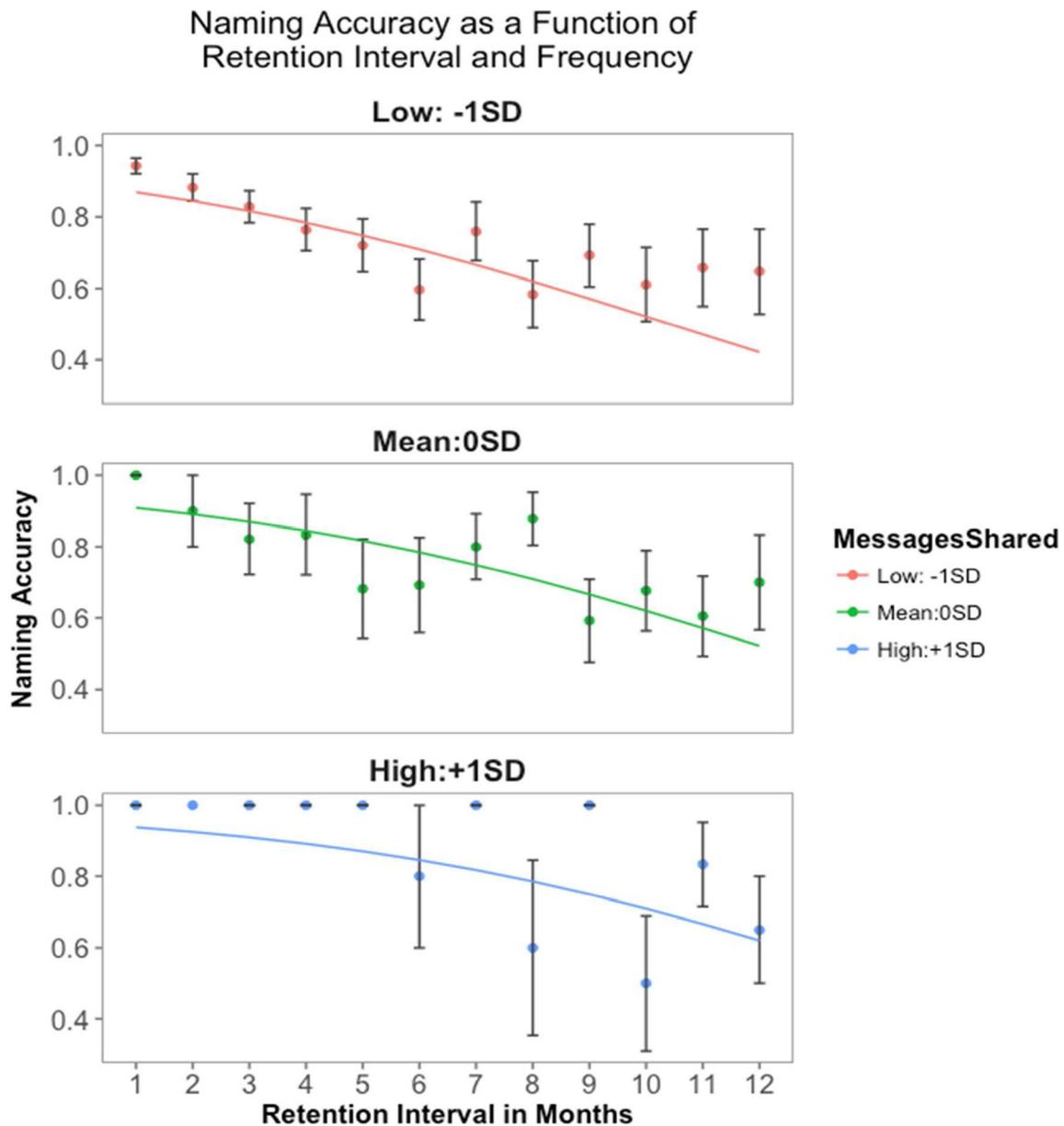
Fig. 4 Plot of recipient name accuracy as a function of retention interval (month) and person-centered frequency (number of messages), with predicted best-fitting lines. Error bars represent standard errors

Table 3 Model estimates for recipient name accuracy

Term	Predictor(s)	Odds Ratio	CI	Std. Error	z Value	p
Fixed parts	Intercept	3.23	(1.35, 7.54)	0.43	2.74	.006
	Month	0.82	(0.77, 0.86)	0.03	− 7.21	< .001
	Mean contact frequency (scaled)	1.13	(0.97, 1.32)	0.06	1.58	.113
	Person-centered contact frequency (scaled)	2.32	(1.43, 4.05)	1.29	3.19	.001
Random parts	Subject	0.59				
	Item	0.26				

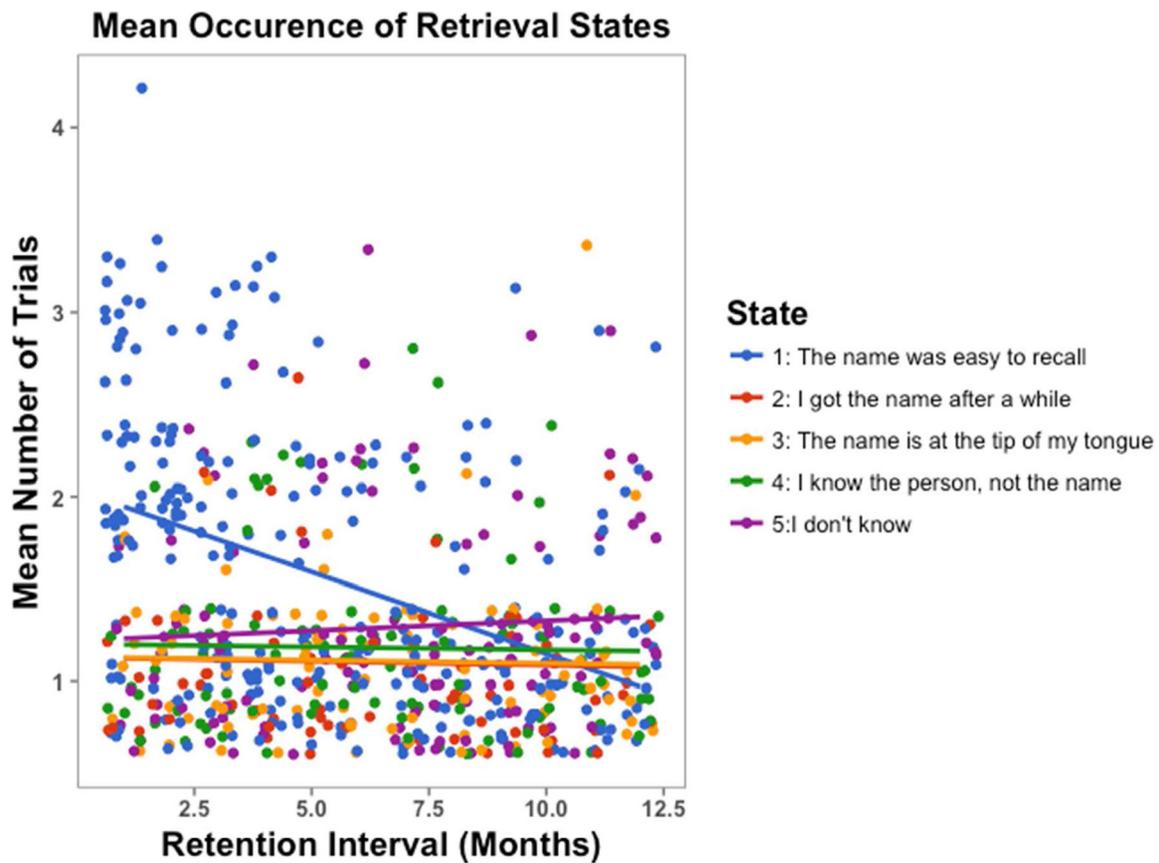Recipient name accuracy ~ Predictors + Random effects of subject and item

Fig. 5 Plot of retrieval states for the e-mail sentence cue as a function of retention interval in months

incrementally in order to evaluate whether each predictor improved the model. Model fit was assessed using chi-square tests on the log-likelihood values to compare models.

Table 4 displays model estimates and coefficients for the final best-fitting model. Model comparisons indicated that retention interval predicted the magnitude of temporal dating errors, but we also observed a significant interaction

between retention interval and the type of temporal dating error, $t = 3.02$, $p = .002$. Figure 6 displays the distribution of temporal dating errors as a function of retention interval and type of temporal dating error. Our results indicate that temporal dating accuracy was greater for more recent e-mails and that temporal dating errors increased linearly with time when the participant made a telescoping error

Table 4 Model estimates for temporal dating errors

Term	Predictor(s)	Estimate	Std. Error	t Value	p
Fixed parts	Intercept	1.343	0.169	7.93	< .001
	Month	0.058	0.032	1.80	.075
	Type of dating error (Telescoping)	− 0.517	0.159	− 3.25	.001
	Sentiment	0.033	0.011	2.96	.003
	Month × Type of Dating Errors (Telescoping)	0.079	0.026	3.02	.003
	Month × Sentiment	− 0.003	0.002	− 1.65	.100
	Type of Dating Errors (Telescoping) × Sentiment	− 0.042	0.011	− 3.74	< .001
	Three-way interaction	0.006	0.002	3.31	.001
Random parts	Subject	0.37			
	Item	0.00			
	Month	0.12			
	Residual	1.29			

Temporal distance ~ Predictors + Random effects of subject and item. All estimates are in reference to the time-expansion error group.
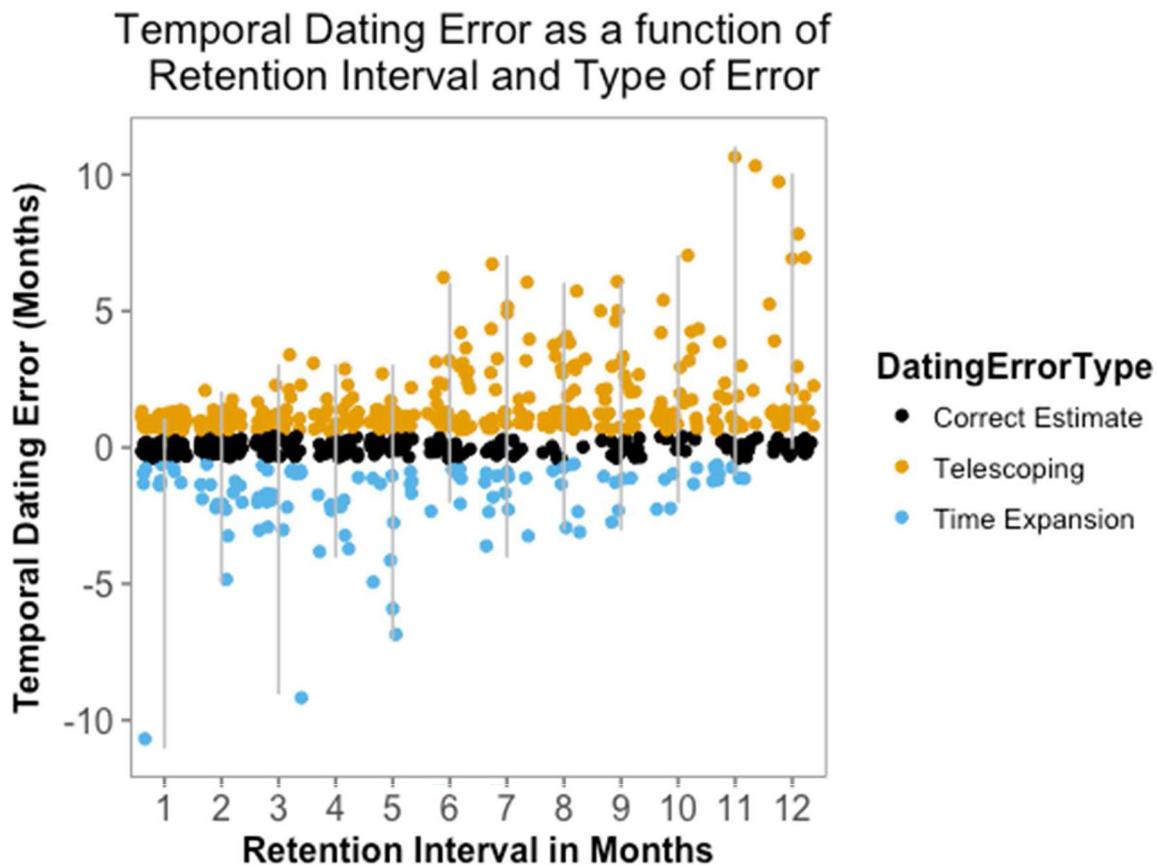
Fig. 6 Distribution of dating errors for the e-mail as a function of retention interval and type of temporal dating error (telescoping, time expansion, or correct estimate)

($\beta = 0.16$, $p < .001$)—that is, guessed a month that was more recent that the actual month in which the e-mail had been written. This slope did not have the same magnitude for time-expansion errors; that is, when participants guessed a month that was farther away than the actual month, the trend in the data was much weaker ($\beta = -0.10$, $p = .048$). Furthermore, sentiment score qualified this two-way interaction, $t = 3.31$, $p = .001$, and the resulting model was significantly better than the previous model ($\Delta AIC = 11.1$, $p = .001$). Figure 7 displays the three-way interaction between retention interval, sentiment score, and type of dating error, as predicted by the best-fitting model.[1] Specific comparisons indicated that when the e-mail contained more sentiments, people were more likely to

make a telescoping error, as compared to a time-expansion error.

Discussion

Methodological innovation

Previous research on autobiographical recall has shown that memory for events depends on the retention interval (Wagenaar, 1986), frequency of exposure to the episode (Bluck & Li, 2001), and the vividness of the event (Sheldon & Levine, 2013). The present study replicated several of these patterns and extends them to the domain of e-mail, suggesting that memory for online communications behaves similarly to autobiographical memory for everyday life episodes as measured in laboratory settings. In addition, we provided a novel, Web-based empirical approach to studying memory for e-mail communications through automated text-based analysis. To our knowledge, this approach is the first to demonstrate that automated text-based analysis of online communications can be used to study autobiographical memory processes.

The present findings also address some important methodological limitations of previous studies. For example, past

---

[1] To effectively display the three-way relationship between retention interval, type of dating error, and sentiment score, a categorical measure of sentiment score was computed, consisting of three levels: "low," "mean," and "high." On the basis of the mean sentiment score across all trials, all trials below one standard deviation of the mean sentiment score were categorized as "low," and all trials above one standard deviation were categorized as "high." The remaining trials were categorized as "mean" items. The mixed-effect model analyses used the actual sentiment score as an interval-type predictor—these classifications into high versus low sentiment were made only post hoc, for display purposes only.
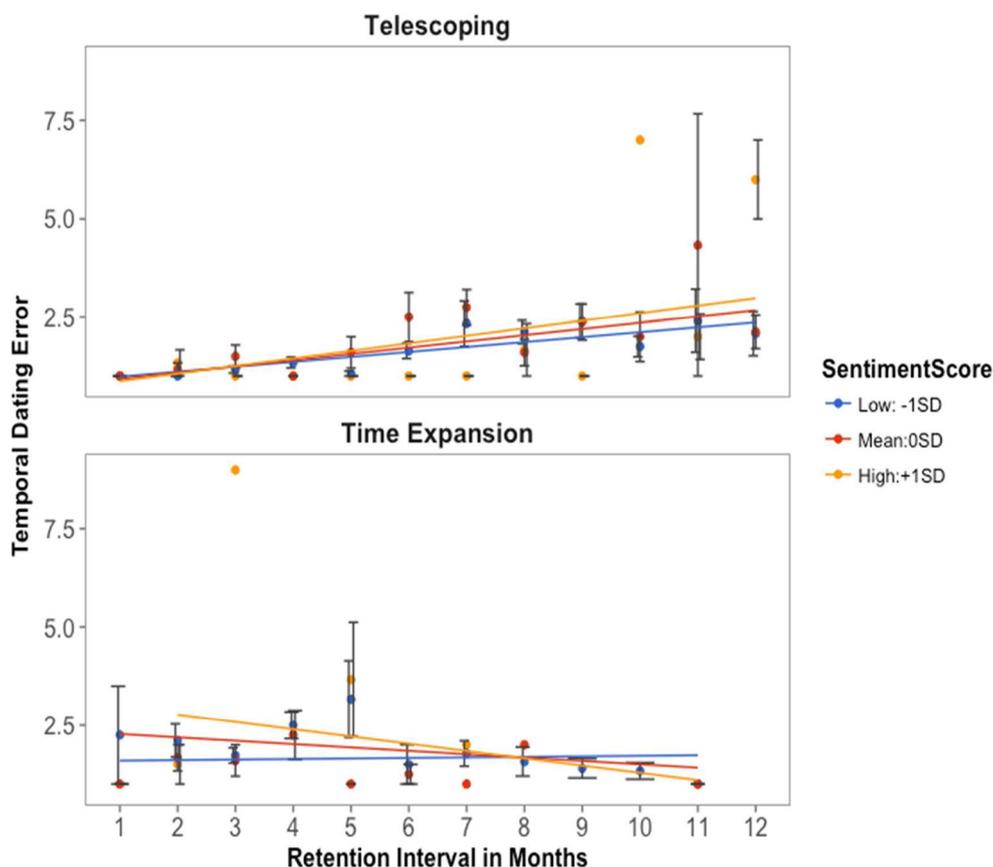
Fig. 7 Plot of the magnitude of dating error for the e-mail as a function of retention interval, sentiment score, and type of temporal dating error (telescoping or time expansion)

work on autobiographical memory has used structured or semistructured questionnaires (Levine et al., 2002), diary studies (Thompson, 1982; Wagenaar, 1986), or wearable cameras (Berry et al., 2007; Dennis et al., 2017), all of which have the potential to introduce demand characteristics into the study of autobiographical memory, since participants are asked to actively remember or record events in these studies. The present study overcomes this bias by generating sentence cues from preexisting e-mail communications in order to probe participants' memory for the e-mail. This approach reduces the likelihood of rehearsal of the memory episode—that is, the e-mail itself—since participants were not aware at the time of writing the e-mails that their words would subsequently be used as memory targets.

Another important difference between the present approach and previous work on autobiographical memory was the generation and use of distinctive and personalized sentence cues for memory testing through automated text-based analysis. Previous work in this domain has used standardized prompts that probed participants to recall specific memories that fit those prompts. For example, as we previously discussed,

Piolino, Desgranges, Benali, and Eustache (2002) had participants recall episodes regarding professional meetings, a trip, a family event, and so forth. These cues typically evoke highly salient memories within individuals but may not reflect their memory for everyday events that are not necessarily as salient. The present approach allows for the possibility of testing memory for everyday life episodes in the form of e-mail communications, and thus provides the opportunity to access a broader set of autobiographical events, without actively having participants provide details for these events. Furthermore, given that e-mail data are time-stamped and tagged, they provide unbiased information about the veracity of the details of the e-mail—for example, the name of the recipient, the exact date on which the e-mail was sent, and so forth. To our knowledge, this is the first automated approach to studying autobiographical memory for everyday communications.

An additional important contribution of this study was the ability to study recall for personally familiar names through automated text-based analysis of e-mail communications. We have provided a novel method of obtaining a large set of personally familiar proper names and studying the impact of

contact frequency on recipient name accuracy. The retrieval of proper names—specifically, the names of familiar people and acquaintances—is especially affected by age (Burke, MacKay, Worthley, & Wade, 1991) and some types of aphasia (Cohen & Burke, 1993), and forgetting the names of familiar people is a commonly reported cognitive complaint among older adults (Ossher, Flegal, & Lustig, 2013). However, previous studies on proper names have tested recall for famous faces, fictional characters, and such (Condret-Santi et al., 2014; Juncos-Rabadán et al., 2013; Middleton & Schwartz, 2013); personally familiar names have remained understudied because of the lack of reliable, long-term estimates for variables that may influence their recall (Brédart et al., 2005). The present approach overcomes some of these problems and provides estimates for contact frequency and retention interval from the last e-mail with the recipient, spread across a year of e-mail-based communications, allowing us to systematically study recall for personally familiar names of people. Additionally, the use of personally familiar names in this approach allows for testing autobiographical memory and familiar name retrieval patterns in a cross-cultural context. For example, names of famous people and common objects may largely vary in different cultures, but our methodology opens up the possibility of using personalized corpora (i.e., e-mail), to generate stimuli for experimental tasks, and address concerns about stimuli familiarity. On the basis of our sample of participants recruited in India, we have provided evidence for findings that are consistent with several laboratory-based findings about autobiographical memory for personally familiar names and also introduced a powerful tool for studying cross-cultural memory processes. We now discuss the specific findings from this study.

Influence of retention interval and contact frequency

The results from the present study suggest that memory for the name of e-mail recipients decreases with longer retention intervals. This decline in recipient memory was also reflected in retrieval states, such that participants were less likely to choose the option "the name was easy to recall" for e-mails at farther-away time points. This finding is consistent with several past studies that have shown that autobiographical memory for events declines over retention intervals when participants recall events after recording them in diaries (Thompson et al., 1988; Wagenaar, 1986), answer questions about life events across time periods (Levine et al., 2002), and report events in response to specific prompts (Piolino et al., 2002). Thus, it appears that e-mail communications show a pattern of memory decline similar to that for real-life episodes, suggesting that e-mails function as real-life episodes for people who actively use e-mail. This finding not only replicates previous work but also extends it to a new, online setting.

The retention function observed in this study had a form consistent with previously described retention functions, such as the exponential (Rubin, 1982) and power (Wagenaar, 1986) functions. Indeed, a power function successfully fit our data, with $\gamma = -.16$, and provided a better fit than the exponential function (see Fig. 8). These results indicate that the time course of forgetting observed in e-mail communications is very similar to the time course observed in other forms of autobiographical memory. Several computational accounts have been proposed to explain the factors that influence the retention function (Crowder, 1976; Howard & Kahana, 2002; Murdock, 1960). More recently, Brown, Neath, and Chater
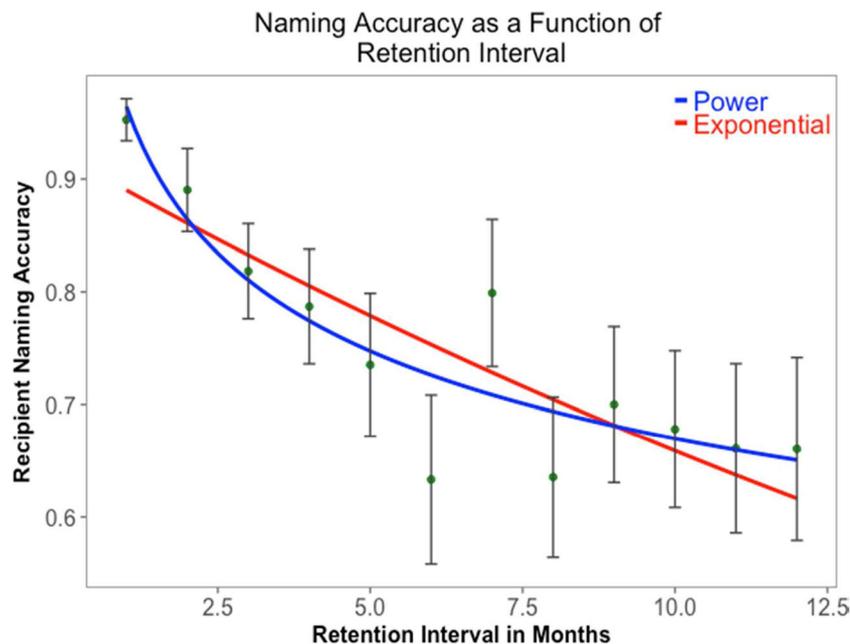

Fig. 8 Plot of recipient name accuracy as a function of retention interval, overlaid with power and exponential retention functions

(2007) proposed a temporal distinctiveness model of memory retrieval, according to which items or events are represented in memory along a temporal dimension, and memory loss is a function of how easily an event or item can be distinguished from its psychological neighbors along this continuum. The notion here is that recent events occupy less confusable locations along the temporal dimension and are thus more retrievable than other items. The results of the present study nicely dovetail with the predictions from this computational account. However, it is important to acknowledge that the interpretation that follows is entirely post hoc, and demands further inquiry. We propose that recent e-mails are remembered better due to their better discriminability, whereas recall for remote e-mails is lower, likely due to interference from other older e-mails. However, other nontemporal factors may impact item discriminability, such as semantics and source information (Brown et al., 2007; Polyn, Norman, & Kahana, 2009). In the present study, the distinctiveness of a particular e-mail may be a function of the personalized sentence cue, which potentially serves to "isolate" the e-mail from its other psychological neighbors along the same temporal dimension, thus facilitating recall. Further, the extent to which a particular e-mail is remembered may also depend upon the social cluster to which the recipient belongs, the number of other e-mails sent during the same time period etc.

We also observed an effect of contact frequency on memory for recipients, such that participants were more likely to recall the name of the e-mail recipient if they had shared greater number of e-mails over the year with the recipient. Importantly, this effect of contact frequency did not interact with retention interval. Thus, retention interval and contact frequency produced additive influences on the memory for the recipient name, a finding that has been previously reported by Anderson and Schooler (1991) in their analysis of the *New York Times* and e-mail communications, as discussed earlier. The effect of frequency of exposure on autobiographical memory has been previously studied (Bluck & Li, 2001), although it has involved repeated exposure to the same event. In the present study, however, participants only viewed the sentence cue once, and no e-mail recipient was repeated within any participant, although the participant likely communicated with the recipient over the course of the year. Importantly, the sentence cue was chosen from the participant's *last* e-mail communication with a particular recipient. The present findings suggest that frequency of e-mailing with a recipient is an important predictor of subsequent retrieval of their name, in response to the sentence cue. Frequency of name use has also been identified as a critical variable in lexical retrieval of proper names (Brédart et al., 2005; Burke, MacKay, Worthley, & Wade, 1991), suggesting that retrieval of the name of the e-mail recipient from the sentence cue in the present study possibly involves complex interactions between lexical units for the name and the memory trace

for the e-mail episode. Furthermore, this finding further supports the notion that although e-mails are represented in a temporal continuum (Brown et al., 2007), other factors, such as frequency of interaction, may increase an e-mail's distinctiveness and facilitate subsequent recall.

An alternative hypothesis for the effect of contact frequency on recipient name recall may suggest that participants were using base-rate information in the task. Specifically, if participants only e-mail a limited number of people and have a general sense of their e-mailing frequency with the recipients, they may employ statistical guesswork in order to arrive at the correct answer. For example, if a participant only e-mails Fred, Sarah, and Bob, and 80% of those e-mails are to Bob, then the participant might just respond "Bob" when presented with an e-mail that they do not remember sending. There are a few reasons why this may not have been true in the present study. First, the false alarm rate in this study was very low; that is, participants entered an incorrect name on only 5.6% of the total trials initially, and on 5.5% of the total trials after the letter hint had been provided. This indicates that on most trials, participants either entered the correct answer or refrained from providing any response. This argues against the possibility that participants were using base-rate information to guess recipient names, else we would have observed very high rates of reporting frequently contacted recipients on incorrect trials.

Due to the nature of our protocol, we did not have access to the participants' complete address books at the end of the study. Thus, to test this alternative hypothesis, we were only able to analyze a small subset of the trials (16%) on which the participant entered an incorrect name. Specifically, we analyzed trials on which the participant had incorrectly entered the name of a recipient who was also part of the study. In this way, we were able to examine whether the incorrect name that a participant responded with was indeed among the frequently contacted recipients for that participant. Our results indicated that for all such trials, the total number of e-mails sent to these recipients fell within two standard deviations of the mean number of e-mails sent by the participant. Furthermore, only 40% of the recipients fell outside one standard deviation of the mean for the participant, indicating that when participants did enter an incorrect name, their guesses were not driven by frequently contacted recipients. It is also noteworthy that there was no systematic pattern in the errors that participants made; that is, no particular name was repeated more than once by any participant. Additionally, on the basis of the distribution of total messages sent ($M = 10.59$, $SD = 22.94$) in our sample, we examined the effects of retention interval and contact frequency after excluding trials for frequently contacted recipients (> 1 SD)—that is, we excluded all trials on which the number of total messages sent to the recipient exceeded 33.53. The effect of retention interval ($p < .001$) and person-centered contact frequency ($p = .025$) persisted in these

analyses, providing stronger support for the hypothesis that contact frequency was indeed influencing recall, and that participants were not simply employing statistical guesswork in this task.

Finally, as we discussed before, accuracy for recalling the name of the e-mail recipient was very strongly correlated with vividness ratings and reported retrieval states in our data, further arguing against the possibility that participants were guessing during the task. Instead, it appears that participants were indeed more likely to recall the name of the recipient from the sentence cue if the recipients had been contacted more frequently. We believe that this finding reflects an increase in distinctiveness and salience for the sentence cue, due to frequent communication with the recipient, which may have more easily evoked the memory of the e-mail episode for the participant.

Temporal dating errors

We were also interested in the nature of temporal dating errors made by participants in this study. Our results indicate that the magnitude of dating error for when the e-mail was written increases with longer retention intervals, particularly when the participant makes a telescoping error—that is, dates an e-mail more recent than it actually is. Importantly, we found a dissociation between telescoping and time-expansion errors, such that whereas telescoping errors reliably increase linearly over time, time-expansion errors showed a very weak trend in the opposite direction. Furthermore, the vividness of an e-mail conversation was negatively correlated with telescoping errors ($r = -.15$, $p = .003$), but not with time-expansion errors ($r = -.01$, $p = .888$), suggesting that telescoping errors have a different pattern from time-expansion errors. This finding is consistent with previous research on the temporal dating of memory events. For example, Thompson, Skowronski, and Lee (1988) found that whereas telescoping errors increased linearly over time, there was no such reliable effect for time-expansion errors. Furthermore, they also showed that this pattern could not entirely be attributed to guessing by participants: That is, it was not the case that participants were simply more likely to guess that an event had occurred at a moderate retention interval when they were unable to date the event (i.e., regression to the mean). They argued that if this were indeed the case, they should also have observed time-expansion errors at recent retention intervals, which should have systematically decreased over time, and this did not occur. Rubin and Baddeley (1989) suggested that this dissociation occurs due to greater retention for recent events and the fact that intrusions can occur from outside the testing period, but these intrusions cannot come from events that have not occurred, thus creating unequal distributions for telescoping and time-expansion errors. We replicated this pattern of temporal dating errors within the domain of e-mail

communications, again suggesting that memory for e-mail communications is similar to memory for real-life episodes.

Furthermore, we also found an impact of sentiment on the type and magnitude of temporal dating errors, such that participants were more likely to date an event as more recent than its actual date (telescope) if the e-mail had higher sentiments, and that they were less likely to date such e-mails as more remote than their actual date (time expansion). We believe that the sentiment score coded by our algorithm might reflect the emotional salience for the event, and previous research has shown that emotional intensity can enhance autobiographical memory for the event (Talarico et al., 2004). Thus, to the extent that sentiment increases the distinctiveness of an e-mail, people are more likely to guess that the e-mail was more recent than it actually was. To our knowledge, this has been the first study to show that sentiments in e-mails can be informative about, and sensitive to, temporal memory for the e-mail. Interestingly, we did not see a relationship between frequency of e-mailing behavior and the magnitude of dating errors, suggesting that frequency of communication with a person does not influence the ability to correctly date a memory episode.

Limitations

The present study also had some important limitations. First, the study was based on e-mail communications, which represent a restricted sample of everyday-life episodes, and this may also point to a selection bias, since we cannot assume that e-mails are the primary medium of communication or a complete record of everyday events. Furthermore, we recruited participants for this study on university e-mail portals, and the participants who volunteered to take part consented to share their e-mail content, so that our encrypted program could generate test items for them. Additionally, due to technical complications, we also excluded some participants who did not use a Gmail account or did not have an adequate number of e-mails to generate a sufficient number of test items. These methods may have affected the representativeness of our sample, since only a small proportion of university students and faculty actively use their e-mail in India, and an even smaller fraction would be willing to share their e-mail content with a third-party program. These constraints may have influenced the results of this study. However, a restricted sample would have possibly led to a smaller spread in accuracy scores and the magnitude of dating errors, as compared to the true population, which would have possibly attenuated the relationships identified in this study. Thus, it is possible that even stronger relationships could be found in a more representative sample from the population.

Second, the findings from this study about memory for the e-mail reflect the last e-mail interaction between a participant and an e-mail recipient. However, since participants retrieved

the names of personally familiar people in this study, it is possible that the participant in fact had interacted with the recipient through another medium of communication, which might enhance the memory trace for the name of the e-mail recipient. The present study does not have any systematic method of addressing this concern; future studies that make use of this approach would need to validate this method and control for other variables, such as offline contact frequency and the recency of offline communication. However, to the extent that our findings reflect memory for the specific episode of writing the e-mail itself, we provide strong evidence that memory for online communications behaves similarly to vivid real-life episodes.

Finally, the features identified and extracted from e-mails in this study were based on those in previous work (Hangal et al., 2011), and the specific weights assigned to the features were empirically derived, on the basis of several rounds of iterative pilot testing. However, a more fine-grained analysis of memorable features would be an important next step to validate these measures. Specifically, other than the high levels of recall observed in the study (which provides evidence for face validity), we did not have any systematic way of assessing whether the features and sentiments picked up by our algorithm indeed reflect highly salient, memorable cues for our participants. A simple way to address this would be to ask participants to rate how well the cue represents the contents of the specific e-mail at the end of the experiment in follow-up studies. On the basis of anecdotal evidence from previous work with MUSE, and the post-experiment survey, participants generally found the study informative and thought most of the cues were indeed relevant. Furthermore, as is shown in Fig. 3, only 11.5% of the sentence cues were rated as being "vague" by participants, suggesting that the algorithm detected memorable sentence cues for the majority of the trials. However, a systematic study of the effectiveness of the cues generated and how well they corresponded to real-world interpretations of memorability and high affect would certainly be more informative and conclusive.

Ethics and privacy

The present study on e-mail communications is part of a movement in psychological science toward using powerful, automated online tools to conduct psychological research (Allen & Roberts, 2010). Internet-based studies provide the opportunity to not only access larger and more demographically diverse samples (Naglieri et al., 2004), but also use technological innovation to collect important data on a larger set of variables than would be possible in a laboratory setting (Hoerger & Currell, 2012; Kraut et al., 2004). However, online studies also raise important ethical questions about data confidentiality, thorough informed consent, and cross-validity. In addition to these concerns

about online research, the present study analyzes personal e-mail text to study autobiographical memory processes, and it is important to consider the potential risks and ethics associated with fully automated research tools (Emery, 2014). Given the sensitive nature of personal e-mail, we were extremely cautious in our data collection methods and ensured that e-mail confidentiality was maintained at every stage of the experiment. Complying with our IRB protocol, all e-mail content except the sentence cues and answers was deleted immediately after the sentence cues had been generated. Furthermore, we did not store any of the sentence cues after the experiment was completed. Additionally, to address any concerns regarding storing IP addresses (Nosek, Banaji, & Greenwald, 2002) and hacking of personal information, we also offered the option of downloading the test program and participating in the study "offline" and e-mailing us the results, although none of the participants in the present study used this option. We are not aware of any potential security breaches, and no participant expressed concerns about their confidentiality being compromised, during or after the experiment.

It is also important to ensure a thorough informed consent process in online studies (Emery, 2014), especially because the researcher may not be immediately available in the event of a crisis (Buchanan & Williams, 2010; Nosek et al., 2002). Through a detailed informed consent process prior to the study, we ensured that participants were fully aware of the potential risks of participating in a fully automated online research study using their e-mail. The consent form outlined potential risks and benefits and encouraged them to contact the research team in case of any concerns. Participants were also free to quit at any point during the study by simply closing the test window, and could choose to not provide any demographic information they did not wish to disclose, consistent with American Psychological Association (APA) guidelines (APA, 2002; Emery, 2014). Finally, we also provided individualized feedback on study performance through an error-validation phase and also included a postexperiment questionnaire in which participants could ask questions or provide comments and feedback to the research team, as recommended by other online researchers (Emery, 2014). Despite undertaking these measures to ensure that the present study closely followed APA guidelines for "best practices" for conducting psychological research, some aspects of the present study were not optimal—for example, the lack of debriefing procedures for participants who might leave the study prematurely, not being able to intervene in the case of misunderstandings during the study, clarifying individualized feedback, and so forth. Future studies that make use of this method should ensure that the methods are appropriately modified to address these concerns.

Future directions

There are several important future directions for this work. Although there have been several informative studies on online communications on public listservs (Park & Conway, 2017) and active online communities (Tan, Niculae, Danescu-Niculescu-Mizil, & Lee, 2016), our automated, Web-based approach enables researchers to analyze participants' e-mail communications and create highly personalized stimuli with precise timestamps. We have made this methodology available to future researchers, so that it can be further extended to other behavioral paradigms and answer more focused questions about autobiographical memory, proper-name retrieval and language use. Quantifying the distinctiveness of an e-mail along temporal and nontemporal dimensions and understanding the factors that influence its retention would be an important next step. Furthermore, questions about the structure and scale of long-term memory (Moreton & Ward, 2010) could be answered through free recall paradigms based on easily verifiable, time-stamped e-mail events. Additionally, MUSE's algorithms for fetching and cleaning e-mail content, and using named entity recognition and other machine learning techniques to browse e-mail archives have been further developed to enable access and analysis of e-mails through ePADD (Schneider et al., 2017). ePADD empowers researchers to use preexisting e-mail collections to answer important research questions in a variety of domains.

Conclusion

The present study replicated and extended previous findings about autobiographical memory in an ecologically valid sample of online e-mail communications, spread across a year. The results indicated that memory for e-mail communications is similar to memory for real-life events, as studied in laboratory experiments. We also introduce a novel, Web-based empirical approach to studying memory processes that might identify new patterns and provide useful applications to the study of autobiographical memory and personally familiar names in a cross-cultural, online setting.

References

Allen, P. J., & Roberts, L. D. (2010). The ethics of outsourcing online survey research. *International Journal of Technoethics*, *1*, 35–48.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct (Online resource). Retrieved December 7, 2010, from https://www.apa.org/ethics/code/.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, … D. Tapias (Eds.), Proceedings of LREC 2010 (pp. 2200–2204). Luxembourg City: European Language Resources Association.

Bates, D. M., & Sarkar, D. (2005). The lme4 library (Software library). Retrieved from https://CRAN.R-project.org/package=lme4

Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., … 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*, 582–601.

Betz, A. L., & Skowronski, J. J. (1997). Self-events and other-events: Temporal dating and event memory. *Memory & Cognition*, *25*, 701–714.

Bluck, S., & Li, K. Z. H. (2001). Predicting memory completeness and accuracy: Emotion and exposure in repeated autobiographical recall. *Applied Cognitive Psychology*, *15*, 145–158. https://doi.org/10.1002/1099-0720(200103/04)15:2<145::AID-ACP693>3.0.CO;2-T

Borrini, G., Dall'Ora, P., Della Sala, S., Marinelli, L., & Spinnler, H. (1989). Autobiographical memory. Sensitivity to age and education of a standardized enquiry. *Psychological Medicine*, *19*, 215–224.

Brédart, S., Brennen, T., Delchambre, M., McNeill, A., & Burton, A. M. (2005). Naming very familiar people: When retrieving names is faster than retrieving semantic biographical information. *British Journal of Psychology*, *96*, 205–214.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Buchanan, T., & Williams, J. E. (2010). Ethical issues in psychological research on the Internet. In S. D. Gosling & J. A. Johnson (Eds.), Advanced methods for conducting online behavioral research (pp. 255–271). Washington, DC: American Psychological Association.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults?. *Journal of Memory and Language*, *30*, 542–579.

Cohen, G., & Burke, D. M. (1993). Memory for proper names: A review. *Memory*, *1*, 249–263.

Condret-Santi, V., Barragan-Jason, G., Valton, L., Denuelle, M., Curot, J., Nespoulous, J. L., & Barbeau, E. J. (2014). Object and proper name retrieval in temporal lobe epilepsy: A study of difficulties and latencies. *Epilepsy Research*, *108*, 1825–1838.

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

Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society: Series B*, *38*, 45–53.

D'Angelo, M. C., & Humphreys, K. R. (2015). Tip-of-the-tongue states reoccur because of implicit learning, but resolving them helps, *Cognition*, *142*, 166–190.

Dennis, S., Yim, H., Sreekumar, V., Evans, N. J., Garrett, P., & Sederberg, P. (2017). A hierarchical Bayesian model of "memory for when" based on experience sampling data. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 295–300). Austin: Cognitive Science Society.

Emery, K. (2014). So you want to do an online study: Ethics considerations and lessons learned. *Ethics & Behavior*, *24*, 293–303.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for*

Computational Linguistics (pp. 363–370). Stroudsburg: Association for Computational Linguistics.

Hangal, S., Chan, P., Lam, M. S., & Heer, J. (2012). Processing email archives in special collections. Paper presented at the Digital Humanities Conference, Hamburg.

Hangal, S., Lam, M. S., & Heer, J. (2011). Muse: Reviving memories using email archives. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (pp. 75–84). New York: ACM Press.

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., … Wood, K. (2006). SenseCam: A retrospective memory aid. In International Conference on Ubiquitous Computing (pp. 177–193). Berlin: Springer.

Hoerger, M., & Currell, C. (2012). Ethical issues in Internet research. In S. J. Knapp, M. C. Gottlieb, M. M. Handelsman, & L. D. VandeCreek (Eds.), APA handbook of ethics in psychology Vol. 2. Practice, teaching, and research (pp. 385–400). Washington, DC: American Psychological Association. https://doi.org/10.1037/13272-018

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299. https://doi.org/10.1006/jmps.2001.1388

Juncos-Rabadán, O., Facal, D., Lojo-Seoane, C., & Pereiro, A. X. (2013). Does tip-of-the-tongue for proper names discriminate amnestic mild cognitive impairment?. *International Psychogeriatrics*, *25*, 627–634.

Kampf, M., Nachson, I., & Babkoff, H. (2002). A serial test of the laterality of familiar face recognition. *Brain and Cognition*, *50*, 35–50.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences*, *101*, 3310–3315. https://doi.org/10.1073/pnas.0306408101

Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The autobiographical memory interview: a new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of Clinical and Experimental Neuropsychology*, *11*, 724–744.

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, *59*, 105–117. https://doi.org/10.1037/0003-066x.59.2.105

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

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, *12*, 271–295. https://doi.org/10.1017/S0305000900006449

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: The posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, *104*, 667–676.

Middleton, E. L., & Schwartz, M. F. (2013). Learning to fail in aphasia: An investigation of error learning in naming. *Journal of Speech, Language, and Hearing Research*, *56*, 1287–1297.

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, *17*, 510–515. https://doi.org/10.3758/PBR.17.3.510

Murdock, B. B., Jr. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology*, *60*, 222–234. https://doi.org/10.1037/h0045145

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, *59*, 150–162. https://doi.org/10.1037/0003-066X.59.3.150

Nagpal, A., Hangal, S., Joyee, R. R., & Lam, M. S. (2012). Friends, romans, countrymen: lend me your URLs. using social chatter to personalize web search. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (pp. 461–470). New York: ACM Press.

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*, *112*, 11078–11083. https://doi.org/10.1073/pnas.1507104112

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*, 161–176.

Ossher, L., Flegal, K. E., & Lustig, C. (2013). Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition*, *20*, 220–242.

Palombo, D. J., Williams, L. J., Abdi, H., & Levine, B. (2013). The survey of autobiographical memory (SAM): A novel measure of trait mnemonics in everyday life. *Cortex*, *49*, 1526–1540.

Park, A., & Conway, M. (2017). Longitudinal changes in psychological states in online health community members: Understanding the long-term effects of participating in an online depression community. *Journal of Medical Internet Research*, *19*, e71. https://doi.org/10.2196/jmir.6826

Piolino, P., Desgranges, B., Benali, K., & Eustache, F. (2002). Episodic and semantic remote autobiographical memory in ageing. *Memory*, *10*, 239–257.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*, 129–156. https://doi.org/10.1037/a0014420

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org.

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior*, *21*, 21–38. https://doi.org/10.1016/S0022-5371(82)90423-6

Rubin, D. C., & Baddeley, A. D. (1989). Telescoping is not time compression: A model. *Memory & Cognition*, *17*, 653–661.

Rubin, D. C., & Siegler, I. C. (2004). Facets of personality and the phenomenology of autobiographical memory. *Applied Cognitive Psychology*, *18*, 913–930. https://doi.org/10.1002/acp.1038

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Schneider, J., Chan, P., Edwards, G., & Hangal, S. (2017). ePADD: Computational analysis software facilitating screening, browsing, and access for historically and culturally valuable email collections. *D-Lib Magazine*, *23*(5/6). https://doi.org/10.1045/may2017-schneider

Sergent, J. (1986). Microgenesis of face perception. In Aspects of face processing (pp. 17–33). Berlin: Springer.

Sheldon, S., & Levine, B. (2013). Same as it ever was: Vividness modulates the similarities and differences between the neural networks that support retrieving remote and recent autobiographical memories. *NeuroImage*, *83*, 880–891.

Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y., Maeda, Y., Matsue, Y., … Kawashima, R. (2006). Cortical mechanisms of person representation: Recognition of famous and personally familiar names. *NeuroImage*, *31*, 853–860.

Talarico, J. M., LaBar, K. S., & Rubin, D. C. (2004). Emotional intensity predicts autobiographical memory experience. *Memory & Cognition*, *32*, 1118–1132. https://doi.org/10.3758/BF03196886

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In Proceedings of the 25th

International Conference on World Wide Web (pp. 613–624). Geneva: International World Wide Web Conferences Steering Committee.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54.

Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, *10*, 324–332.

Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1988). Telescoping in dating naturally occurring events. *Memory & Cognition*, *16*, 461–468. https://doi.org/10.3758/BF03214227

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

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, *48*, 63–68.