What’s-Its-Name: Cognitive Experiments with Life-logs (CELL), A New Approach to Measure Recall of Personally Familiar Names

Abhilasha A. Kumar  
Ashoka University  
Haryana, India  
abhilasha.kumar@ashoka.edu.in

Allyson C. Rosen  
Stanford University  
California, USA  
rosena@stanford.edu

Vihari Piratla  
Amuse Labs, Dharwad,  
Karnataka, India  
viharipiratla@gmail.com

Sudheendra Hangal  
Ashoka University  
Haryana, India  
hangal@ashoka.edu.in

ABSTRACT

Everyone has had the experience of forgetting the name of a person, place or movie and remembering it later. While these personally familiar names pervade our lives, the mechanisms to explain their recall have not been properly studied. Cognitive Experiments on Life-Logs (CELL) is a scalable new approach to measure recall of personally familiar proper names using computerized text-based analysis of email archives. Emails today are storehouses of time-stamped, personally relevant information, which can be leveraged to study memory processes and proper name recall. In this paradigm, 44 participants viewed sentences automatically drawn from a year of their sent email, and recalled the name of the email recipient. Regression analyses revealed that accuracy in familiar name recall and vividness of the memory declined with the age of the email, but increased with greater frequency of interaction with the person. These findings suggest that the recall of familiar names is mediated by both delay and frequency of name use, and that CELL can be applied as an ecologically valid web-based measure to study memory and language processes using existing life-logs.

Author Keywords

Proper-name recall; life-logging; autobiographical memory

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

INTRODUCTION

Recalling names of personally familiar people and places is central to everyday life, whether it involves greeting an acquaintance, retelling the adventures of a trip to a friend, or simply revisiting a familiar place. This ability to recall proper names in everyday life contexts decreases with age and with degenerative brain function, but the mechanisms of this memory decline are not properly understood. A common difficulty that arises in this study of personally familiar proper-names is that they are often associative to multiple life contexts and cannot be localized to a unique encoding context. A systematic study of proper-names thus requires the identification of specific variables of interest from these life contexts, which can directly be linked to cognitive processes related to memory. We propose a new, automated and scalable approach in this direction, Cognitive Experiments with Life-Logs (CELL), using email archives. The extensive use of email makes it a thriving storehouse of personally important information. These life-logs represent time-stamped events and memories, which can be analyzed using text-based analysis and machine-learning algorithms. Our method identifies potentially memorable sentences from people’s sent emails and presents them as stimuli in a name retrieval task, to study familiar name recall in a naturalistic setting. Through this email-based paradigm, we demonstrate that the memory for names of familiar people declines with the age of the email and lower frequency of interaction. Our algorithm also provides data on over 80 variables of interest including reaction time, vividness, gradient of recall and sentiments in email, which can be further analyzed to study familiar name recall.

Studies on proper-name recall and the closely related phenomenon of the tip-of-the-tongue (TOT) state have been based on names of famous faces, places, fictional characters or films [9, 19, 23]. Personally familiar proper-names, on the other hand, are said to be associated with real-life events. Therefore, their recall would involve retrieving the name and the associated episodic memory, unlike the process used for recalling famous names, which are represented in a largely semantic context [35]. For episodic memories, the encoding specificity principle [37] posits that recall is improved when information available at encoding is available at retrieval, and therefore, invoking identical contexts during retrieval should facilitate recall. If emails are considered as “episodes” in some sense, their text could be used to simulate episodic context for the recall of proper names present in emails. Mantyla (1986) has argued that compatibility and distinctiveness of cues determine recall, and our paradigm made use of such compatible and distinctive cues to construct test items for familiar name recall. As delay since encoding impacts memory accuracy and vividness, in our previous study [27], we explored this phenomenon by testing people’s memory for personally familiar proper names from their email archive. In the study, participants viewed sentences picked from their email archive, where a critical proper name had been
omitted. (For example, ‘ _ _ _ _ _ _ _ _ sat on a wall’). Participants were required to fill in the blank with the correct proper name (‘Humpty Dumpty’ in the present example), where these names were 40 memorable and frequently occurring proper names in the participant’s one-year archive, and included names of people, places, books etc. We were able to detect a significant memory decline in healthy adults, over the period of a year. An important insight that emerged from the study was that the memory for familiar proper-names was strongly affected by time delay.

The present study focused on specifically testing the recall for names of people, to understand the effects of delay and frequency. The retrieval of familiar proper names, and within them, the names of familiar people, is especially affected by age and some cases of aphasia [6], and older adults commonly complain about forgetting names of people [24]. A systematic study of the factors that impact the memory for names of familiar people, with stimuli drawn from an ecologically valid setting could enhance our understanding of the mental processes involved in name retrieval. Furthermore, neuropsychological studies have suggested that name retrieval deficits may be useful in distinguishing between amnestic mild cognitive impairment (aMCI) and healthy aging adults [25], especially in light of the fact that people with MCI are at a high risk of developing Alzheimer’s disease [21].

Burke (1991) demonstrated that the memory difficulties associated with proper-names is more pronounced specifically for names of acquaintances and famous people when compared to memory for other names of films, places etc. among older adults. Cohen and Faulkner (1986) explained this differential memory for different types of proper names through a ‘gradient of recall’ based on meaningfulness, in that names of people are arbitrary, lack substitutes and would thus be more difficult to recall than other types of names. There are several factors that may contribute to the recall of names of people. Familiarity and recency of usage have been found to play a key role in determining recall performance for names of public figures and uncommon objects among older and younger adults [26]. Frequency of name use is also understood to be a critical factor in mediating recall [2], although estimates of familiar name frequency have been difficult to obtain due to
lack of systematic data on name use [10]. Given that CELL’s text-based analysis has been developed to sift through time-stamped emails for proper names and derive several metrics of name use, we also studied the effect of frequency of name use on recall through this paradigm. In the study, a two-sentence prompt picked from the participant’s sent email was shown, and participants were required to name the recipient of the email. We explored the effect of time delay and frequency of use on familiar name recall. With respect to delay, we predicted that in the manner of the memory for real-life episodes, the accuracy and vividness for the memory for the email, but also the names associated with it would decline over time. This hypothesis is consistent with studies on proper-name recall, where recent exposure to a target name has been shown to reduce incidences of retrieval failure, and also tip-of-the-tongue states [5]. In other studies, recall for infrequent words was found to be lower than for higher frequency words [5], as was the incidence of TOTs [38]. Overall, we predicted that the recall proportion for familiar names would increase with greater frequency of use, and decrease with time delay. Our results provide a proof of concept that existing life-logs can be leveraged to study cognitive processes related to memory and language.

METHODS

Participants
Participants were 37 young (M=24.13, SD=2.7) and 7 middle-aged adults (M = 49, SD = 8.40) recruited from Ashoka University and its extended network of faculty, parents and founders in India. In order to minimize the complexity of coding in the study, only participants with Gmail email archives were allowed to participate. The study was approved by the Ashoka Institutional Review Board and participants provided informed consent online.

Overview
The task was designed such that the participant retrieved the name of a personally familiar person from memory based on a sentence prompt from their own email (Figure 1). Upon viewing the prompt, participants entered the name of the email recipient in a text box. To generate these sentence stimuli, we applied text-based analysis following a set of rules to the last one year of the participant’s email. These rules identified stimuli that would be memorable to the participant and also required the participant to retrieve a unique proper name from memory. To systematically understand the effect of time delay on recall, we identified memorable stimuli from every month from a year-old email archive of the participant. The computerized generation of these test items required extracting certain parameters from the emails without human intervention and devising a scoring algorithm, which is discussed in a subsequent section. Participants completed the test phase where they were asked to type the name of the email recipient, and also specify the nature and extent of their recall. For each test item, a phonemic first-letter hint appeared on the test screen after their initial response, so that the participant could use the hint, if they were unable to retrieve the name earlier. For each test item, participants also specified the vividness of the email conversation and the month in which they wrote the email. After completing this test phase, participants completed an additional error judgment phase, where they assessed their errors and report them as valid or invalid. These metacognitive judgments were useful in understanding whether CELL’s text analysis truly captured personally relevant information, and in measuring accuracy, in cases where the algorithm failed to detect correct answers due to minor spelling errors or the use of other names of the recipient.

Procedures
After informed consent, the participants’ experience was as follows. Participants waited while their email archive and demographic information was evaluated to determine whether they would qualify for the study. We imposed a screening criteria on the participants’ email archive, such that participants were required to have sent roughly 10-12 emails in each month of the past year, in order to generate sufficient number of test items for each participant. Upon clearing the screening, they were then given a code that enabled them to return to the study to participate at a time when they could spend approximately 30-45 minutes of focused time to the task. Following a brief explanation about the questions and a demonstration of the task, they entered the first test phase. During this phase, participants completed all the test items and for each test item, they answered a series of questions about their responses. After the first test phase, they began a second, error judgment phase during which they were shown both the errors they made and the correct answer and they made judgments about the errors.

Stimuli

Data Preparation
Before text-based analysis could be performed, data from the participant’s email archives was fetched and prepared for analysis. This involved the participant logging into their email, after which all the sent emails in the past year were fetched to an encrypted, secure server housed in the Department of Computer Science at Ashoka University. This excluded email attachments, formatting of text, images and other non-text data. Further, quoted and forwarded sections of the email were removed from the archive, and since our paradigm currently focused only on Gmail accounts that have fixed formats for quoting and forwarding, this task was accomplished robustly by identifying fixed templates. We also performed entity resolution to match different email addresses belonging to the same person, in order to generate a contact address book. This step was crucial to the test phase, where the participant recalled the name of the email recipient, which was then matched to the name in the address book. After questions and answers were prepared by the program, the remaining email archive was deleted from the system. The
time required in generating test items and questions was included with the time taken to fetch and clean the data, so that there was no delay during the test phase.

Generating Relevant Proper Names of Persons
Names of people were identified from the contact address book, and scored based on their frequency of interaction. Adjustments were made to exclude generic mailing lists, and people who had been emailed only once, as recall for these recipients was expected to be poor. For every month of the year, names of people who were last emailed in that month were chosen, and persons with whom communication had been concentrated over a specific time period (2 months) around that month were scored higher, in order to identify names of time-specific acquaintances as potential test items. To avoid confusing test items with more than one correct answer, we only considered emails sent to one recipient in constructing sentence prompts for that recipient. However, to evaluate whether a recipient was a good candidate for a given time interval, we tracked all communication with the contact, in single and group emails and factored it into the score of the contact. An important constraint in identifying these proper-names was that they should have been accompanied by a reasonable sentence prompt that would enable the participant to guess these names. The method of generating the sentence prompts is discussed in the next section, but the final test items were chosen by comparing the combined score of the proper name as well as the accompanying sentence prompt for all possible name-sentence prompt pairs and choosing, maximally, the three highest scoring pairs as the tests item for that month. There were instances wherein the participant email archive fell short of three recipients last contacted in a particular month, in which case as many as available, but at most three such contacts was chosen for the sentence-prompt generation phase.

Generating Sentence Prompts
After identifying personally memorable names of people, the challenge lay in constructing a two-sentence prompt that would provide adequate context and enable the participant to accurately retrieve a unique proper name. The rationale for choosing a two-sentence prompt was based on our previous work and pilot testing, where participants reported that one sentence often did not succeed in triggering the memory of the email. We reasoned that two-sentence prompts not only gave the participant adequate context to follow a previously written line of thought, but also eliminated one-sentence emails that weren’t memorable as such. To construct these sentence prompts, we programmed a tokenizer that retrieved all sentences from the participant’s last email written to the selected recipient, along with a set of features. These features were then analyzed and scored using empirically-derived metrics of memorability based on our previous work [17].

The features for each sentence prompt were generated by the tokenizer so that they could be used to identify memorable sentences. The tokenizer tracked only sent emails excluding forwarded email, to ensure that participants had mentally processed the text at some time. Our previous work and pilot testing revealed that some sentences, such as ‘Hi, it was great meeting you’ and ‘Looking forward to meeting you’ were generic enough to be answered with any proper name. Through multiple rounds of pilot testing and iterations across different email accounts, we identified features that made a context distinctive and memorable and, therefore, required the retrieval of a unique recipient. For example, memorable sentences often contained emotional content, indicated by sentiment words such as ‘happy’, ‘sad’ etc. or emoticons, such as 😊, 😒 etc. Question marks and exclamation points also indicated memorable sentences. Sentences that contained other proper names were also more memorable than generic sentences. These names were identified using the ‘email: Process, Appraise, Discover, and Deliver’ (ePADD) Named Entity Recognizer (NER). Common features of proper names are that they are often mentioned in initial capitals, are names of geographical locations or famous personalities or are preceded by salutations or greetings. Adjustments were made to filter out word variations, capitalized pronouns and strings of capitals longer than 15 characters, as these were typically not names. We also excluded common internet abbreviations, such as BTW, FYI, PFA etc., and eliminated sentences that contained the participant’s own name or the recipient’s name, as such prompts would defeat the purpose of the name retrieval task.

A particular set of features tracked the number of sentences, the length of sentences and the number of sent emails in a thread. In this step, we also performed a sentence-to-sentence matching, to identify and exclude exactly same sentences used across emails, since they did not serve as particularly unique and distinctive cues. In the scoring phase, we assigned low scores to sentence-prompts that spanned beyond a specified range of characters and lines, to avoid extremely long or short sentence prompts and maintain uniformity among test items across and within participants. This policy also controlled for cases in which participants had simply pasted long text from other sources, which could include a lengthy passage from a book or article, or a fragment of code. We also assigned higher scores to sentence-prompts that were part of a longer thread in which the participant had emailed more than once, or contained sentiment words, question marks and exclamation points. All of these features relevant to the sentence prompt, the recipient and the email containing the prompt were assigned weights and numerically coded into an aggregate final score. Finally, for each month, at most three sentence prompts with the highest final score, indicating a personally significant name and memory were chosen as the test item(s) for that month.
Hints During Test Phase

Two types of hints were provided to participants. Initially, the number of letters in each part of the name was displayed using dashes (eg. For ‘Humpty Dumpty’: __ __ __ __ __ __ , and also in text form (eg. two words [5 letters, 5 letters]). After responding to the sentence prompt, by either typing the name, or characterizing their extent of recall, participants were provided the initials of the name as a second hint (eg. H __ __ __ D __ __ __ ). The purpose of providing hints was to ensure that the participant did not substitute a related but different name. This way, the participant had additional information about the proper name to be able to eliminate related names that may have come to mind.

Metacognitive Judgments

Participants gave qualitative and quantitative judgments about their responses after each test item, specifically on the extent of their recall, vividness levels and recency estimates. For specifying the extent of recall, participants responded to the statement, ‘Tell us about your recollection’, by choosing one of the five options, presented in the increasing order of recall strength: ‘the name was easy to recall’, ‘I recalled the name after a while’, ‘the name is at the tip of my tongue!’, ‘I know the person, but not the name’ and ‘I don’t know’. This metric is derived from other commonly used standards for measuring the nature of recall, where the response is characterized as Know- TOT – Don’t Know [33, 31, 10]. Based on feedback and observations from our pilot work, we further parsed this response to capture the entire spectrum of responses that may occur while retrieving the name of a person. For vividness judgments, participants answered the question, ‘How vividly do you remember this conversation?’ (on a visual scale displaying numbers from 1 to 10) whereas for time judgments, participants answered, ‘Approximately when did you write this email?’ and they provided the month and year of the email. Our previous study [27] had demonstrated that people were highly accurate of estimating the month of the email, independent of their recall of the proper name.

RESULTS

Error Validation

The first step in our analyses was to evaluate whether our automated approach could be relied to produce valid test items, and eliminate the need for human intervention, especially since emails are personal archives. After the first test phase, participants marked their errors as valid or invalid in the error judgment phase. Items were marked as correct if the name typed by the participant produced an exact match with the name picked from the contact address book. Therefore, we considered all correct items as valid. Among the error types (1: I feel like I should have remembered this name, 2: My answer is essentially correct, 3: I recognized the context, but not the person, 4: I have trouble remembering this name, 5: It was a very vague/generic sentence), we considered Error Type 2 and 5 as invalid errors. Figure 2 displays the percentage of these error types across all participants. Error Type 2 is relatively high (35%) because people often typed in the first name, or nickname, or misspelled the name, in which case their memory for the proper name was intact but the program would have marked their response as incorrect. A personal evaluation by one of the authors (A.K.) revealed that such judgments were accurate, and they were thus counted towards the final score of correct responses per participant. In Error Type 5, participants indicated that the sentence prompt did not provide enough context for them to have guessed the email recipient. Such errors would be dramatically higher if items were picked at random and our algorithm was not applied, but with our algorithm, these errors were 22% of the total errors. Pilot testing suggested that such errors occurred for test items such as, ‘Let’s meet on Tuesday at 10 PM’ or ‘I am working as a consultant at XYZ’, which failed to provide a unique encoding context, and participants’ judgments were usually accurate about these items. Since our protocol did not store any of the sentence prompts in the actual study, a direct review of the test items was not possible, in which case the judgments were assumed to be fair, and the corresponding test items were eliminated from all further analyses. In addition, all items that the participants skipped were also excluded from all further analyses.

Delay Effects

Accuracy over Time

Two separate types of analyses were performed to assess the effect of the age of the email (in days) on the recall of the email recipient name. The program recorded delay in terms of the exact number of days from the day of the test. However, to obtain more stable estimates for the effect of delay, test items were grouped into four intervals spanning three months each (30 days). Several aggregate analyses...
Figure 3. Plot of mean proportion of correct responses per quarter. Effects were significant at the 0.01 level (2-tailed), and a declining trend was observed. Error bars represent confidence intervals.

were thus of the form of a repeated-measures ANOVA with the delay variable spanning the most recent quarter (1) to the most remote quarter (4), from the day of the test. To measure overall accuracy, we combined items marked as correct by the program with those that were marked as “essentially correct” by the participant, and calculated proportion of correct responses for each quarter. A repeated-measures ANOVA, with Greenhouse-Geisser corrected (\(\varepsilon = 0.73\)) degrees of freedom revealed a significant decline in proper name recall over the course of the year, \(F (2.18, 84.96) = 7.532, p = 0.001\) (Figure 3).

Bonferroni post-hoc comparisons revealed a strong recency effect, with a significant drop-off in name recall after the first quarter. Since delay was measured on a continuous scale, we also performed regression analyses to be able to predict recall over time, which will be discussed in a subsequent section.

Reaction Time over Time

Reaction time is a powerful measure of proper-name recall, often used in lexical decision and word retrieval tasks. Our program recorded reaction time as the total time spent on each item, which included three separate tasks: name retrieval, recall and vividness judgments, and time judgments. While there may have been individual differences in the time spent on each task for every item, a repeated-measures ANOVA, with Greenhouse-Geisser corrected (\(\varepsilon = 0.74\)) degrees of freedom revealed a significant increase in mean reaction time per quarter, \(F (2.21, 86.066) = 3.55, p = 0.029\) (Figure 4), suggesting that it took participants longer to respond to older emails. Bonferroni post-hoc tests showed that there was a significant rise in reaction time between the first and third quarter, as well as the first and fourth quarter.

Vividness over Time

If emails resemble time-stamped episodic memories, we hypothesized that vividness for the email conversations would systematically decline over the period of a year. Participants rated their vividness for the email on a scale of 1 to 10 (in response to, “How vividly do you remember this conversation?”). We calculated mean vividness ratings for each participant per quarter, and a repeated-measures ANOVA with Huynh-Feldt corrected (\(\varepsilon = 0.87\)) degrees of freedom revealed a systematic decline in vividness across quarters, \(F (2.62, 102.39) = 18.55, p < 0.001\) (Figure 5).

Extent of Recall over Time

After attempting to recall the name of the email recipient, participants were asked to provide qualitative responses about their extent of recall (the answer to, ‘Tell us more about your recollection’, with anchors as in Table 1). Since the anchors were ordered from 1 to 5, 1 representing maximum recall (‘the name was easy to recall’) and 5 representing minimum or no recall (‘I don’t know’), we examined whether the extent of recall correlated with delay. We computed correlations between the recall type and days of delay for each participant and tested whether these values differed from 0. 41 out of 44 participants demonstrated an expected positive correlation over time, with a significant average effect, \(t (43) = 7.829, p < 0.001\).
Metacognitive Test-Item Judgments

**Extent of Recall:** “Tell us about your recollection.”

1. The name was easy to recall
2. I recalled the name after a while
3. The name is at the tip of my tongue!
4. I know the person, but not the name
5. I don’t know

**Vividness:** “How vividly do you remember this specific conversation?”
10 options on a visual scale from 1-10, with 1 indicating no memory, 5: fair idea and 10: strong memory

**Time Judgment:** “Approximately when do you think this sentence was written?”
13 options, from the name of the current month and year, backwards in time e.g. June 2016, May 2016 etc.
These are followed by the option ‘I have no idea’

Table 1. Three judgments were collected after every response for every test item. Participants first tried recalling the name, and then entered their responses for extent of recall, vividness and time judgments.

and a repeated-measures ANOVA also revealed a significant decrease in extent of recall, \( F(3,117) = 12.364, p < 0.001 \), confirming the hypothesis that the extent of recall fades over time, from maximum recall (1) to little or no recall (5).

**Time Judgments**

For each test item, participants provided their estimate of the month in which they wrote the email. Even though the program records the age of the email in days since the test, we computed the month of the email to compare it with the participants’ estimate on the same scale. We performed separate correlation analyses for correct and incorrect responses, because our earlier study had revealed that people were surprisingly accurate about the month of the email, even for those items where they couldn’t guess the proper name. We wanted to understand if the pattern would recur when they were asked to guess the name of the email recipient, as opposed to filling the blank. Items for which the participant did not guess the month at all were removed from further analyses. Figure 6 shows an overall scatter-plot between the guessed month and actual month, and also indicates a positive slope for the fitted line for both correct and incorrect items. For correct items, all 44 participants showed a high positive correlation between month guessed and actual month, with a significant average effect, \( t (43) = 55.92, p < 0.001 \). For incorrect items, 28 out of 44 participants showed a positive correlation between the month guessed and the actual month, with a significant average positive effect, \( t (27) = 11.05, p < 0.001 \). We also observed significant positive partial correlations between month guessed and actual month, after controlling for participant ID, for correct items \( (r = 0.887, p < 0.001) \) and more crucially, for incorrect items \( (r = 0.809, p < 0.001) \). This finding is important, in that it demonstrates that participants retained some information about the email, despite not being able to name its recipient, suggesting differences in recall pertaining to items and their contexts. The implications of this finding are discussed subsequently.

**Frequency Effects**

A critical factor in choosing test items for this study was the frequency and nature of interaction with the email recipient. For every month, the algorithm picked contacts that had last been contacted in that month, and then applied a set of rules to generate reasonable sentence prompts for them. Further, contacts that had been emailed several times in the same month were preferred over contacts emailed only once or twice. This resulted in a large range of frequencies in the final test items for the total number of messages sent to the

Figure 6. Overall scatter plots of guessed month and actual month for correct (left) and incorrect items (right) respectively. Both plots show a positive correlation.
email recipients by the participant over the year. This presented an opportunity to understand how frequency of interaction affected the recall for names of familiar people. We also performed regression analyses to model this frequency effect, which is discussed in a subsequent section.

**Accuracy, Extent of Recall and Frequency**

We first performed correlations between total messages sent to the email recipient over the year and accuracy, where 40 out of 44 participants demonstrated the expected positive correlation, and a significant average effect, \( t(43)=6.621, p < 0.001 \). We also observed a significant positive partial correlation between accuracy and total messages, controlling for the participant ID, \( r = 0.126, p < 0.001 \), suggesting that more interactions with a person resulted in greater memory for their name.

We also computed correlations between total messages sent to the email recipient, and the extent of recall (Table 1). 39 out of 44 participants demonstrated the expected negative correlation, and a significant average effect, \( t (43) = -7.636, p < 0.001 \). A significant negative partial correlation between the total messages shared and the extent of recall, controlling for the participant ID was also observed, \( r = -0.108, p = 0.002 \) suggesting that greater extent of recall for the name (level 1) is related to higher number of sent messages with the person.

The best fitted model for this paradigm is also shown in Figure 8 and the models is ess ken Information. These patterns suggest a nearly linear decline in -d

**Vividness and Frequency**

Correlations between vividness ratings and total messages sent revealed that 36 out of 44 participants demonstrated the expected positive correlation, and a significant average effect, \( t(43) = 6.108, p < 0.001 \). A significant positive partial correlation between vividness and frequency, after controlling for participant ID was also observed, \( r = 0.095, p = 0.006 \), suggesting that higher frequency of interactions is also related to greater vividness for the email conversation.

**A Regression Model for Familiar Name Recall**

There were two key variables of interest in this paradigm: the days of delay and the frequency of interaction (calculated by the total number of messages sent to the email recipient over a year). Since both variables were continuous, and their overall effect on familiar name recall was consistent with our hypothesis, we were interested in understanding whether they would predict the accuracy of name recall. Mixed effects models were used as the primary statistical procedure for this analysis, implemented through the statistical program R. Since the response variable, accuracy was dichotomous (0/1), and each participant responded to several test items, each of which had values for accuracy, days of delay and frequency of interaction, we performed a repeated-measures binary logistic regression. Three separate mixed effects logit models were fitted to the data: the null model, the days of delay model and the complete regression model. Fixed effects inclusion in the model was evaluated on the basis of likelihood ratio tests, thus including only effects which significantly increased the model’s goodness-of-fit. We included a by-participant random effect, and considered scaled days of delay and scaled frequency of interaction as fixed effects. For the sake of simplicity, we report only the parameters of the final, best-fitting models with the lowest Aiken Information Criteria (AIC) and log-likelihood scores, together with their significance level based on Satterthwaite's degrees of freedom approximation in the lme4 R package (version 3.3.0 (2016-05-03)).

A summary of the fixed effects of each of the models is reported in Table 2. The best fitted model for this paradigm of familiar name recall included scaled days of delay and frequency of interaction as fixed factors (Model 3). The Hosmer and Lemeshow goodness of fit (GOF) test indicated no significant difference between observed and fitted values, \( \chi^2 (8) = 9.971, p = 0.26 \), suggesting a good model fit. The mean accuracy was 75.86% (range 38.46 - 95.65, SD = 14.54). Analysis of accuracy revealed that days of delay significantly predicted an accurate response (Wald \( Z = -5.18, p < 0.001 \)), and frequency also significantly predicted an accurate response (Wald \( Z = 3.63, p < 0.001 \)). The predicted probabilities of accuracy w.r.t. days of delay and frequency are shown in Figure 7. The plot of fixed effects for each participant is also shown in Figure 8 and Figure 9. These patterns suggest a nearly linear decline in predicted accuracy with days of delay, and an exponential
growth in predicted accuracy with increasing frequency of interaction, which is consistent with our hypothesis and previous results.

DISCUSSION

Two important findings emerge from this study: the memory for names of familiar people most strongly decreases with days of delay since last contact with the person, and also increases with greater frequency of interaction with the person. This approach is the first to demonstrate that text-based analysis of life-log data can be used to study cognitive processes related to proper-name retrieval. Based on an algorithmically chosen set of stimuli, participants were not only able to retrieve the name of the email recipient, but also provide fairly accurate time and error judgments. Our previous study [27] was conducted via Craigslist in the United States of America, in which we obtained a powerful of time delay. The present study was conducted at Ashoka University in India, where we not only replicated previous findings pertaining to time delay and time judgments, but were also able to find an effect of frequency on familiar name recall. This method can thus measure the factors affecting proper name recall cross-culturally, and this study demonstrates that CELL is a feasible, ecologically valid and culturally sensitive approach that can be applied to further study cognitive functioning, without human intervention.

We now discuss some specific findings and their implications. Participants’ memory for name of familiar

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coefficient</th>
<th>SE</th>
<th>odds ratio</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.54</td>
<td>0.15</td>
<td>4.64</td>
<td>10.03</td>
<td>&lt;2e-16 ***</td>
</tr>
<tr>
<td>delay</td>
<td>-0.47</td>
<td>0.09</td>
<td>0.62</td>
<td>-5.18</td>
<td>2.2e-07 ***</td>
</tr>
<tr>
<td>frequency</td>
<td>1.42</td>
<td>0.39</td>
<td>4.14</td>
<td>3.63</td>
<td>0.000279 ***</td>
</tr>
</tbody>
</table>

Table 2. Statistics of three mixed effects logit models fitted to recall data

Figure 8. Plot of predicted probabilities of accuracy with days of delay for each participant. Days of delay have been scaled, with center at 154 and SD = 105.96.
people declined over the year. We also observed a very strong recency effect in our analysis of accuracy over time, in that the greatest drop-off occurred after the first quarter of the year. The principle of recency holds for autobiographical memory processes [13, 28], and studies have demonstrated that the greatest amount of forgetting occurs in the weeks following an event [29, 30], which in this particular case is sending the email. Vividness of autobiographical memories is also an important measure that is said to account for differences in hippocampal activation between recent and remote memories [32], and we found a significant decline in vividness of the email over the period of the year. Further, participants were surprisingly accurate of the month of writing the email even for incorrect items, suggesting that memory for the context (the sentences) as opposed to that for the item (name of the email recipient) relies on different recall processes, evidence for which has been found in several item-context studies [36]. These findings motivate the discussion of familiar-name recall towards a finer understanding of the context in which it occurs. Semantic dementia patients have been found to recognize their own objects but not similar objects belonging to others [34, 1, 14], recognize famous places they have visited more than those they have not, and identify relatives over famous faces [34, 16], suggesting that the names they are able recall or recognize are embedded in powerful, familiar and personal contexts that facilitate recall. In brain pathologies like semantic dementia, where there is a breakdown of abstract knowledge, patients tend to rely on meaning derived from personal, everyday experiences [18], necessitating that the mechanisms explaining familiar-name recall incorporate the effects of personal contexts in their narrative. CELL presents the opportunity to systematically identify such contexts, through its time-stamped email data and text analysis, both of which can be used as tools to understand how everyday experiences affect processes of encoding and retrieval.

Another important result from this study was that the recall of names of familiar people increased with greater frequency of interaction with the email recipient. Given that we use familiar names on an everyday basis, to greet, introduce and address acquaintances, and often as attention holders [7], their frequency of use is relatively higher than those of other common names, and would likely be an important determinant of their recall. However, Bredart (2005) [2] points out that frequency of name use is a critical variable that cognitive models of person identification and naming, like the Interaction Activation and Competition (IAC) do not seem to emphasize. Burke and colleagues, on the other hand, have shown that name retrieval difficulties are associated with infrequent, non-recent words [4, 5], and that connections between the lexical and phonological nodes that produce the name are strengthened by higher frequency of use. This finding is of greater importance for familiar name recall, where in as much as recency of exposure is critical, the repeated use of and exposure to familiar names will also affect recall. This study systematically measures the impact of frequency of name use in everyday life, on recall, through rich email archives,

Figure 9. Plot of predicted probabilities of accuracy with frequency of interaction for each participant. Frequency of interaction has been scaled, with center at 10.59 and SD = 22.4.
and empirically demonstrates that frequency is an important determinant of familiar name recall and should be emphasized in cognitive models of name identification.

There are several limitations to this study that need to be considered. This paradigm of testing familiar name recall is built on the premise that emails are similar to memory episodes, and the decline in familiar name recall observed is akin to the fading of episodic memories over time. However, whether emails are a representative sample of everyday life is an open question, although the fact that most participants performed at above-chance level, and were also able to guess when the email was written suggests that emails are indeed memorable parts of their life. However, since the study was advertised on university email portals, participants were not picked at random but volunteered to participate, which may have affected the representativeness of the sample. Further, selecting memorable and valid test items was challenging, and several potential participants who did not have an actively used Gmail account were automatically excluded from the study so that sufficient number of test items could be generated for each time window of the year. Another important consideration in this study pertains to the measure of frequency of name use. In our earlier study [27], we observed the effect of delay, but since an important criteria for choosing the test items was that the proper names chosen occurred repeatedly in the participants’ archives, the effect of frequency on recall was not observed. This study was able to obtain a large range of frequencies, in terms of total messages sent to the email recipient, by not relying on frequency for choosing test items, but instead on the time stamp of last contact with the recipient. While the time stamp was unambiguous and precise, we have no way of knowing when and how often the participant was exposed to these names outside of email. Since our focus was on names of familiar people, it is important to understand that several, if not all of the email recipients chosen could have been people that the participant interacts with on a daily basis in real-life, which would impact their memory for the person’s name.

A common experience associated with proper-name recall is the “tip of the tongue” (TOT) phenomenon, marked by an inability to retrieve the name from memory, but also accompanied by a strong sense of imminent recall [3]. TOTs are typically associated with non-recent, low-frequency words [5], and are explained as a problem of lexical access. In our study, we tested for TOTs through our first metacognitive judgment on the extent of recall, where participants tried recalling the name of the recipient and then answered, ‘Tell us more about your recollection’ (Table 1). Typical studies on TOT make use of the ‘know-don’t know- TOT’ paradigm, which was altered in this study to capture a wider spectrum of responses from the participant. The distribution of responses from all participants is displayed in Figure 10. Overall, participants reported few TOTs (11.91%), and while this could mean a lower incidence of TOTs, it may also be the case that participants did not clearly understand the phrase, in which case they could have chosen option 4 (I know the person, but not the name) or option 5 (I don’t know), or were able to resolve their TOTs after some time, and so they would have chosen option 2 (I recalled the name after a while). The aggregate of options 2, 3 and 4 yields a substantial percentage (34.59%), indicating that participants did experience some form of word retrieval failure, although little can be said about its specific nature. A repeated-measures ANOVA for the occurrence of response types 2, 3 and 4 over quarters was not significant, \( F (3,129) = 1.59, p = 0.194 \), although an interesting trend was observed, in that the incidence of these word-finding failures was largest in the second quarter (Figure 11). Since the focus of this study was to explore overall impacts of frequency and delay
on familiar name recall, a follow-up study to systematically test for TOTs is recommended.

There are several important future directions for this work. This paradigm is the first to demonstrate that text-based analysis of life-logs can be used to study cognitive function, and specifically memory and language processes among people over the web. This study is an important first step in potentially bridging the gap between lab-based studies and everyday functioning. The finding that recency and frequency jointly determine recall for names of familiar people, as is demonstrated by our regression analyses, is interesting and has important implications. Name retrieval difficulties reported by older adults and aphasic patients [6] are often interpreted as problems of lexical access, but the fact that delay emerged as a powerful effect mediating familiar name recall suggests that the everyday experience of recalling names involves both memory and language processes. Neuroimaging studies addressing this particular overlap of memory and language areas could potentially dissociate the brain structures involved at different stage of the name retrieval process and lead to a better understanding of this phenomenon.

A great deal of further work will be required to establish reliable protocols and translate this interesting approach into a clinical measure, although the task could be very rewarding. In cases of brain injury or pathology, it is difficult to obtain measures of earlier mental ability, due to which measures of cognitive decline are less sensitive than they might otherwise be. The names and relevant stimuli extracted through this paradigm could be used to design effective interventions in cases of severe brain injury, as they likely invoke memories from a rich personal past. For cases where people routinely struggle to remember some names, this paradigm could serve as a tool to rehearse and improve everyday recall. Finally, text-based analysis also presents an opportunity to measure sentiment in emails, and a carefully designed protocol could potentially explore the connection between emotion and memory, given that the “emotive quality” of episodes has been suggested as a possible explanation for greater recall of personally experienced events in patients of semantic dementia [16].

CONCLUSION

The objective of this study was to apply a new and scalable scientific method using computerized text-based analysis to identify the factors that affect the recall of personally familiar proper names, which were extracted from existing digital life-logs in the form of emails. Delay since writing the email and the frequency of interaction with the person were found to mediate the recall for the name of the email recipient, suggesting that familiar name recall likely involves both memory and language processes. An additional finding was that participants were able to recall the month in which they wrote the email, even when they were unable to name the recipient, suggesting that the memory for the email context persisted even when the memory for the name faded. These findings call for a finer understanding of familiar name recall and provide a proof of concept for the use of CELL as a method of studying everyday cognitive functioning.

REFERENCES


32. 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.


