Effect Of Phonemic Cuing On Recall Of Personally Relevant Names Derived From Email

Allyson C. Rosen [1], Ankit Mathur [2]; Monica Lam [2]; Sudheendra Hangal [2]
[1]Stanford School of Medicine, [2]Stanford University, Stanford, CA 94305,

Abstract

This is the first study to study memory using email-derived stimuli. Email is a lifelog that can be mined to study cognition in an ecologically valid context. Our laboratory is applying a system that automatically creates stimuli from participants' emails to probe memory for proper names in an autobiographical memory context (MUSE: Hangal, Lam, and Heer, 2011). In this web based study, participants view email messages they previously sent and are asked to insert the proper name that was omitted. The task thus involves recall of a proper name and source memory when the participant matches it to the appropriate email context. We have recently demonstrated that the accuracy of recalling these proper names declines over the course of a year (Hangal, Rosen, Mathur and Lam, 2014). The objective of this current analysis is to test whether a letter cue is helpful for recall and whether the benefit of this phonemic cuing decreases with increasing delays over the course of the year.

Introduction

Everyone has had the experience of struggling to recall a proper name, often when we most need it in everyday life. For example we might need to remember the name of someone we are about to meet for lunch or a place we have to go. Previous studies of proper name recall typically focus on recall of people's names and have involved laboratory experiments, case studies of patients with focal lesions and even invasive brain stimulation (for a review see Hanley, 2014). Ultimately it is important to study the ability to recall personally relevant proper names in their appropriate life context. This study aims to develop an approach to testing people's ability to recall proper names in relation to personally relevant context by selecting email stimuli using computerized text-based analysis (MUSE: Hangal, Lam, and Heer, 2011) and asking people to recall proper names based on the original sentence context. Aside from the benefit of having veridical capture of communication episodes that have been prospectively gathered in an ecologically valid context, emails provide exact information about delay to capture more exact detail about decay over time. The focus of this study will thus be on the impact of delay on measures of memory strength of proper names. Of particular interest was the question of whether access to the proper name was at a lexical level or whether there was more of a problem related to binding to the context. If the problem with name retrieval was lexical, one would expect a letter/phonemic cue to be equally effective throughout the delay. If the process involved binding to the memory context then the cue would be less effective with delay.

Methods

Participants: Forty eight people (average age= 30.2 [range 19-60]) were recruited nationally from Craig's List and paid $10 for participation. All participants needed to be over 18 and to send at least 20 emails per month over the course of a year. To further constrain analyses we limited participation to email accounts from gmail and yahoo.

Stimuli: Email archives were analyzed on a secure, FIPS 140-2 encrypted server at the Stanford Department of Computer Science and 40 sentences were derived as stimuli using a method previously described (Hangal et al., 2011). To ensure that people processed the mails there were several constraints on item selection such as only mails that were in the sent mailbox were included.

Test items were structured so that the task was to type the name that should fill in where the proper name was missing. Items were cued with blank lines substituted for the letters. After 15 seconds with no response, participants saw the first letter of the correct name. (Figure 1). For each test item participants made judgments about their memory for the items (Figure 2) such as how certain, how recent, and how vivid the memory episode was.

Proper names and sentence contexts were selected based on parameters that have been demonstrated to be related to the likelihood that people would be able to recall a unique name. For example it is a challenging problem to generate these stimuli since without these algorithms most sentences would not provide sufficient context ("Hi ___ how are you?"). These features that were used to select items and other descriptors are available for further analyses. Examples are provided in Figure 3.

Procedures: Participants first were screened and then given an example of the task. They went through 40 trials and answered the Test Item Judgments for each trial. Next they were shown their errors and asked to indicate whether those Error Judgments were valid and what the nature of their errors were (e.g. not enough context, a tip of the tongue error).

Figure 1: Proper Name Recall

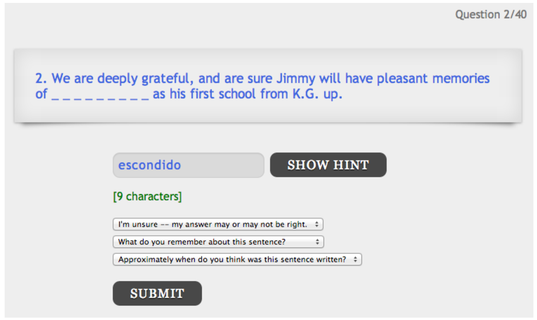


Figure 2: Metacognitive Judgments

Test Item Judgements (Collected after each response)

Certainty: "How sure are you? "
1. I have no idea.
2. I'm unsure -- my answer may or may not be right.
3. I'm fairly sure
4. I'm certain

Vividness: "What do you remember about this sentence?"
1. I don't remember anything about it
2. I can infer the answer, but don't recall this context
3. I only recall the general context, not the message
4. I remember this specific message

Recency: "Approximately when do you think was this sentence written? "
12 options, from the name of the current month, backwards in time.
e.g., September 2013, August 2013, etc.
These are followed by the option ``I have no idea".

Error Judgements: "About this answer…" (Collected after 40 test items completed)
1. I really should have gotten this correct
2. The answer was on the tip of my tongue
3. My answer is essentially correct
4. This is an insignificant detail that I'm unlikely to have remembered
5. The answer is hard to guess... the clue sentence did not provide enough context

Figure 3: Additional Descriptors

Answer term features
- First and last date of usage of term
- Answer appears in any address book?
- Number of messages/threads with the answer
- Monthly histogram of usage

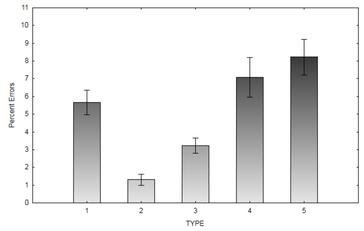
User response features
- Hint used?
- Hint used only for confirmation?
- Milliseconds taken to answer
- Number of messages that a wrong answer occurs in
- Number of messages in which a wrong answer co-occurs with the correct answer

Prompt sentence features:
- Length of the sentence (number of characters)
- Number of named entities in the sentence
- Number of emoticons in the sentence
- Sentence number in the message

Prompt message features
- Age of the message (number of days)
- Sentiment words in message (Categories tracked: superlative, congratulations, grief, anger, confidential, family, religious, love, vacations, racy, emergency, etc.)
- Span of thread containing the message (number of days)
- Number of names in the message
- Number of sentences in the message
- Characters in the subject line
- Answer part of message recipient name?
- Number of recipients

Results

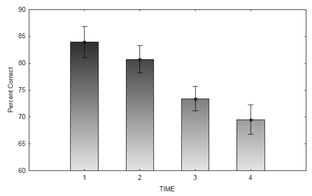
Validation: Evaluation of Error Types



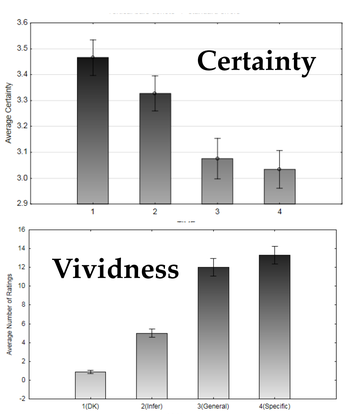
1. I really should have gotten this correct
2. The answer was on the tip of my tongue
3. My answer is essentially correct
4. This is an insignificant detail that I'm unlikely to have remembered
5. The answer is hard to guess... the clue sentence did not provide enough context

Analysis of Error types: Few participants noted trouble with tip of the tongue (Type 2). For items where they reported that their answer was correct they were given credit (e.g. giving Bobby for Robert). This, too, was rare since the answer was constrained based on the number of letters (Type 3). Items were excluded from further analyses where subjects stated they did not have enough context to guess (Type 5) such as a sentence like "Hi, my name is _ _ _". Some of the investigators (S.H. and A.M.) checked several of the response types and in all cases the judgments were reasonable suggesting that this inspection would not been needed in normal participants.

Accuracy Over Three Month Intervals



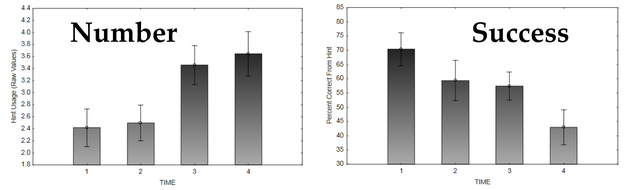
Accuracy (percent correct out of 40 items) were binned into time intervals representing the most recent 3 months (1) to most delayed 3 months (4) over the course of a year. Our hypothesis was that the mental process we were studying involved an event in which there was binding the proper name to its email sentence context so that binning delay based on the email time stamp would be a valid basis for assigning delay. In contrast traditionally, studies of proper names assign delay based on the time window over which the proper name was last in use. The problem with this approach is that there is less control over the consistency in spreading test items over time. An ANOVA of number of events confirmed that there was an unequal number of test items across the 3 month intervals when selecting based on the proper name ($F(3,141)=87.09$, $p < .0001$). Regardless it was still possible to demonstrate a significant decline in accuracy over time ($F(3, 105)=13.969$, $p<.0001$). In contrast when the events were binned by time stamp the number of events in each of the four time bins was not significantly different ($F(3, 141)=1.2103$, $p=.31$) . This analysis also resulted in a significant decline in accuracy over time ($F(3, 141)=7.8336$, $p<.0001$) which is displayed above.



Refer to Figure 2: Certainty: Consistent with the hypothesis that a delay led to a decay in memory integrity rather than a shift in strategy for recall, certainty declined with delay as well ($F(3,135)=13.572$, $p < .0001$)

Refer to Figure 2: Vividness: All means were significantly different based on a Newman Keuls posthoc except 3 and 4 ($F(3,129)=54.09$, $p<.001$).

Hints/ Phonemic Cues



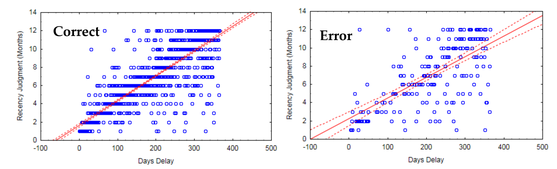
Number: Along with the reduced certainty about their memory, participants asked for hints more frequently ($F(3,141)=4.03$, $p<.01$). Above are the raw number of hints requested which increased by approximately one/time bin at 6 months delay.

Success: Despite the fact that there were more trials with hints requested, the success in those hints in evoking the correct response declined ($F(3,87)=4.39$, $p<.01$).

Latency

Often latency is a more sensitive measure of difficulty than accuracy and we wanted to know if even measuring reaction time in a home setting would show systematic patterns. There was one response that needed to be excluded what was 80,000 milliseconds but otherwise delay was not significant but was similar in direction. ($F(3,117)=1.222$, $p=.3$).

Recency Judgments



In order to test for evidence of source memory, memory for the context surrounding the email event, we evaluated whether there was an association between how accurately people could judge the time of when the email was written and days of delay (Figure 2: Recency). Because the error rates were so low, we performed a fixed effects analysis separately for correct and incorrect responses. The figures above demonstrate that both correct ($r=0.86$) and incorrectly ($r=0.68$) recalled items displayed a correlation between estimated and actual time delay. There were enough items correctly generated to calculate correlation was calculated for each subject and then the correlations were compared to zero. The average correlation was significant ($r=.84$, $t(46)=10.5$, $p<.0001$ with the lowest observed correlation was $r=0.7$.

Conclusions

There are many neuropsychological conditions such as in the case of an acute brain injury where there is no possible way to measure premorbid abilities. Leveraging lifelogs such as email offer this opportunity. This study is the first to apply text based analysis of lifelogs to demonstrate that meaningful and systematic patterns can be abstracted. There key results were:

- Delay-related decline in accuracy, certainty and increased use of hints are consistent with the feasibility of having test items graded in difficulty.
- Phonemic cueing declined in effectiveness with delay and subjects denied tip of the tongue experiences suggesting the problem with accessing proper names not being a function simply of lexical access, but instead of binding word and context.
- Recency judgments in normal subjects were quite robustly accurate even when there was a failure to recall the proper name. This task should be evaluated in clinical populations since it may prove useful.

Acknowledgements

Supported by National Science Foundation

References

Hangal, Sudheendra, Lam, Monica S., Heer, Jeffrey (2011) UIST-2011: Muse: Reviving memories using email archives. Proceedings of the 24th Symposium on User Interface Software and Technology. ACM.

Hangal, Sudheendra, Rosen, A.C., Mathur, A. and Lam, Monica S. (2014) Personalized Memory Testing for Names Using Email Archives Brain KDD Conference Proceedings