Cueing Digital Memory: How and Why do Digital Notes Help Us Remember?

Vaiva Kalnikaitė
The University of Sheffield
Recent Court, 211 Portobello Street
Sheffield S1 4DP, UK
+ 44 (0)114 222 2666
v.kalnikaite@sheffield.ac.uk

Steve Whittaker
The University of Sheffield
Recent Court, 211 Portobello Street
Sheffield S1 4DP, UK
+ 44 (0)114 222 2630
s.whittaker@sheffield.ac.uk

ABSTRACT

People are aware of the fact that their memories are fallible. As a result, they spend significant amounts of time preparing for subsequent memory challenges, e.g. by leaving themselves reminders. Recent findings suggest, however, that people's ability to prepare for subsequent retrieval may not always be effective. This paper looks at the efficacy of memory strategies in the context of digital and paper-based note-taking. Prior research has claimed that (a) notes may not always be useful in promoting later retrieval; (b) taking notes may distract people from effectively processing important information. We examined pen and paper note-taking as well as a new generation digital note-taking device ChittyChatty, finding that notes help memory in two ways. First they provide cues that help people retrieve information that they might otherwise forget. Second the act of taking notes helps people to better focus on incoming information *even if they never later consult these notes*. Finally we found differences between different note-taking strategies. People who take high quality notes remember better than those who focus on exhaustive documentation; taking large volumes of notes decreases the efficiency of retrieval – possibly because it is more time consuming to scan extensive notes to find relevant retrieval cues.

Categories and Subject Descriptors
D.m [**Software**]: Software Psychology.

General Terms
Performance, Design, Experimentation, Human Factors.

Keywords
Memory, Prosthetic memory, Digital memory, Notes, Handwritten notes, Remembering.

Published by the British Computer Society

1. INTRODUCTION AND MAIN QUESTIONS

We are all aware of the fallibility of our unaided organic memories (OMs). In our everyday lives, we often prepare for future memory using a variety of prosthetic memory devices (PMs). We carry PDAs, notepads, diaries and other writing devices to help us remember information that we may need to recall in future [5, 19]. We leave emails in our inboxes or sticky notes and paper files on our desktops when there are outstanding actions associated with these [6, 13, 21, 31, 32]. In the longer term, we create mementos [22] or take photos to trigger memories of events, people and places [10, 16, 24, 26]. Recently there has been much interest in replacing this heterogeneous set of memory devices with so called 'Lifelogging' technologies, as technical developments in capture, storage, and information retrieval now make it possible to record every event we experience [8, 9, 11, 17, 20, 25].

Figure 1. ChittyChatty User Interface

However, the effective use of any PM device requires *strategic planning*, as well as *systematic knowledge* about how our memories work. This knowledge is referred to in the psychology literature as *metamemory* [18]. Metamemory involves knowing for example *what* we are likely to forget and hence what we need to store prosthetically. There is little point in using complex technology to store information that is easily brought to mind, e.g. familiar names, routes or phone numbers [13]. Metamemory also involves knowing *how* to store information; we need to organise it in a way that ensures we can access it when it is needed.

There is, however, evidence that people are not very effective at metamemory strategies associated with stored digital information. First people may store information *that they don't need later*. People prepare for later re-access to web pages by creating complex bookmark hierarchies, but such structures are often not used. For example, 42% of bookmarks are never re-accessed [1, 28]: by far the most common strategy for refinding a webpage is to type its URL [12]. Second, *they may spend time organising information in ways that don't benefit later retrieval*. Studies of email show that some people engage in quite complex filing strategies that are counterproductive – creating folders that contain only one or two messages that make future filing less effective [32]. These results are important because they show that efforts to prepare for retrieval may often be suboptimal.

We examine a common strategic memory behaviour, namely note-taking, as a form of prosthetic memory device. We investigate whether notes *do indeed facilitate* later retrieval of complex conversations. We are all practiced note-takers and it seems intuitively obvious that notes necessarily help memory. One straightforward way that notes can help is to serve as *prosthetic cues*. Looking at a note can directly trigger memory for information such as a name or phone number that is otherwise hard to remember [19, 31]. Note-taking may also facilitate *organic memory directly*. The act of concentrating on key information to compose notes may help OM by causing people to *focus* more on incoming information - even if those notes are never consulted [4, 14, 23].

But these potential benefits of notes are not necessarily guaranteed, and recent research has challenged some of the claims about the utility of notes. As with web bookmarks, we may select the *wrong prosthetic cues* - expending effort noting down information that later turns out to be irrelevant. Various studies have shown that notes do not always successfully cue PM recall especially at longer retention intervals [13]. And notes may turn out to be ineffective if users can no longer remember what they mean [29].

Worse still, taking notes may also *detract* from *organic memory* processes by compromising how we process incoming information. With the demands of complex meetings, we may be so busy trying to record a previous critical point that we miss new important information. This new information may end up being only partly processed and soon forgotten. Thus preparing for future PM retrieval may have *attentional costs* that are detrimental to OM.

Finally we were interested in what motivates note-taking strategies. Are dedicated note-takers people who are aware of the limitations of their organic memory who are interested in offloading retrieval onto PM devices?

We investigate whether, how and why notes might help memory by comparing retrieval for spoken conversations using 2 PM devices, (a) Pen and Paper (PP), and (b) a new generation *digital* note-taking device (ChittyChatty - CC) where digital notes are indexed to a recording of the meeting. We examine how these different PMs are used for retrieving information over different periods of time, and how this compares with unaided OM usage. Further we were interested in what motivates note-taking. Are people who are unconfident about their memory more likely to take notes?

We also wanted to examine different note-taking strategies and how these affect memory. We investigated: (a) *Quality* of PM notes; (b) *Quantity* of PM notes; and looked at how these affect (c) *Memory Performance* (whether using OM, PM or a combination of both) and (d) *Efficiency* of retrieval, i.e. the time taken to remember whether this is with OM or PM. We also conducted an analysis of different note-taking styles, comparing digital and analogue notes along various dimensions.

More specifically we investigated the following research questions:

- *Do notes help overall memory* – regardless of whether people use their notes to answer a question or rely on what they can remember unaided without referring to their notes.

- And if notes do indeed help memory, *how* do they help? Here we distinguish between benefits of notes on PM and OM.

- *PM cueing*: Does note-taking help PM accuracy by generating cues that people use directly to prompt retrieval? Or are such cues ineffective because people expend effort noting information that turns out not to be useful for retrieval?

- *OM effects*: Here we focus on cases where people take notes, but choose not to use them at retrieval – usually because they believe that they can retrieve information correctly unaided. Does careful note-taking *promote OM* accuracy – by causing people to focus more carefully on what was said and hence remember better using OM? Or do notes *detract* from organic remembering? Are attempts to take exhaustive notes counterproductive because users miss much of what was said making them unable to remember little unaided – reducing OM *accuracy*? We call these competing hypotheses *OM focusing* and *OM distraction*.

In addition, we were also interested in the *types* of notes taken and how this affects recall

- *Effects of Note-type on recall Accuracy:* Is it better to take larger numbers of notes (i.e. a large *Quantity* of notes) in order to generate more complete PM cues, at the risk of noting irrelevant information? Or should people employ more concise note-taking strategies that try to focus on more critical information (i.e. high *Quality* notes)?

- *Effects of Note-type on recall Efficiency*: Is it more time-consuming to retrieve information from highly detailed notes? Or do these more exhaustive cues make retrieval more efficient?

We also wanted to explore people's *reasons* for note-taking. How do notes relate to people's *evaluations* of their memory capabilities?

- *Reasons for note-taking:* Are people who are very *confident* about their organic memory less likely to take notes?

Finally we were interested in *general differences* between digital and paper based notes. Do people tend to take the same types of notes in both cases, or are there significant differences between them?

2. EXPERIMENTAL METHOD

We investigated memory using 3 different types of prosthesis: (a) an analogue note-taking system - Pen and Paper (PP), (b) a digital note-taking system called ChittyChatty (CC) which we describe in detail below and (c) no prosthesis (NP) where people rely on unaided memory. Our goal was to test how these different note-taking devices helped users remember everyday conversations. We read a series of conversational stories aloud to users, asking them later to retrieve information about those stories. In the CC and PP conditions, users had a device to help them remember prosthetically by looking at their notes, but they could also choose to rely on unaided OM. In the final NP condition they were not provided with any device support and instead had to rely on OM alone.

2.1 Users

Twenty five users took part (14 women and 11 men, aged 23-55). Users were volunteers consisting of university researchers, administrative and management staff, as well as other professionals from public and private sectors. Users had no prior knowledge of the project or our experimental hypotheses. None of the users had prior experience of using CC, but obviously all had extensive experience with both OM and PP.

2.2 Stories and Test Questions

The 3 stories were intended to simulate real-life conversations between two old friends who had just bumped into one another after a period of several years. The stories contained a mixture of facts and fiction equally distributed within each story. We conducted extensive pilots with the stories, to ensure they could be easily understood, and they did not contain any unfamiliar or unusual terms. User comments indicated that they were enjoyable to listen to, as well as achieving their objective of simulating real-life conversational experiences. The average story time was 3.20 minutes.

An example fragment of one story was the following:

"*Oh, do you remember my older brother, Dave? Let me tell you how he got here. He has loved Def Leppard ever since he was 15 years old and saw them play at the Sheffield Show, Hillsborough Park in 1978. The hair, the tight trousers, the heavy guitars, the thunder of the drums and the screaming vocals. He was particularly entranced with their Yorkshire lyrics. To be honest, he was obsessed. They used to rehearse in some old warehouses and he would hang around outside listening to them tune their guitars. He found it entertaining. When their practice sessions were over, they'd catch their bus home and Dave would pretend he was getting the same bus…In 1979 Def Leppard were one of the biggest rock bands in the country, but then a strange thing happened. A journalist for Sounds magazine wrote that the band had "sold out" to America. Dave wasn't sure what that meant. Like just about every other band, they wanted to be successful in America, but so what? It's not like they had cut their hair, but suits on and started singing mushy ballads. Most of their original fans believed this story and when they played a Reading festival in England, they showered them with bottles. It was another 7 years before their home country would ever really accept them again... Anyway success followed again at the end of eighties, followed by the inevitable decline. The albums began to lose their edge and when Steve, the guitarist died, Dave thought they would pack it all in. But they kept going, keeping the tour bus rolling, last night they came home to Sheffield to play the Arena, and as usual Dave was right at the front going crazy. Suddenly, Joe, the singer spotted Dave in the crowd. He'd recognised Dave after all those years, thought obviously he was a bit fatter and his long hair was thinning a lot. To Dave's complete surprise Joe pulled him out of the crowd and introduced him to the whole arena as Def Leppard biggest fan. They nicknamed him "Mad Dave". Dave raised his arms into the air to bash in his glory and then dived forward back into the crowd. Obviously they didn't fancy a fat, balding and middle-aged rocker landing on their heads. So that's how he got here, Northern General Hospital with crushed ribs, a fractured arm and a broken nose*"

After hearing the story we asked users different recall questions. The above story generated 4 questions: (1) "*Which year did Def Leppard became one of the biggest rock bands in the country?*" (2) "*What was Dave's nickname?*" (3) "*How did the local fans feel about Def Leppard's success in America?*" and *(4)* "*How did the crowd respond to Dave diving onto them*".

2.3 Prostheses

2.3.1 ChittyChatty (CC) - Digital Notes

Fig 1 illustrates the ChittyChatty interface. CC is similar to other note-taking systems such as [2, 27, 29]. Like those systems, it supports memory for conversation using temporal co-indexing [29] of handwritten notes and speech. The main representation is a blank page where users create notes and/or other visual cues while recording a conversation. Users follow their normal practice of taking handwritten notes but each pen stroke is temporally co-indexed with the underlying recorded speech. This allows the notes to be used to access the conversation; when users want to re-access recorded speech, they click on a specific note, and the system begins to replay the speech that was being recorded at the moment that note was taken. In this way the notes serve as a *visual analogue* to the underlying speech, allowing straightforward access to a specific part of the speech. This gives the users a more precise way of accessing a specific part of speech without having to listen to the whole audio again. CC runs on any version of Windows Mobile edition on a PDA, making it portable and easy for taking meeting notes.

2.3.2 Pen and Paper (PP) - Analogue Notes

People were given pen and paper to take notes and they were instructed to take notes as they would normally do to remember a complex verbal materials.

2.3.3 No Prosthesis (NP)

The final condition was NP. In this condition people were not given external memory aids and had to rely on their unaided memory to find the answers to the questions.

2.3.4 Differences in Efficiency and Accuracy between PM Devices and Unaided Memory

Our prosthetic devices have different properties. PP notes are a schematic and incomplete record of what was said, whereas CC offers a verbatim record. Retrieval Efficiency is also different for these devices. Extracting information from PP or written CC notes is efficient because the eye can rapidly scan text to identify information. CC, however, should support reasonably efficient access to the underlying speech record; using handwritten notes or other visual cues should allow users to quickly identify relevant regions of speech to access and listen to.

Of course, both prosthetic devices contrast with unaided memory which is efficient to access but fallible.

2.4 Time: Retention Intervals

The entire experiment consisted of 3 Retention Intervals. The first Retention Interval consisted of an introduction, CC training, exposure to the conversational stories and initial memory testing. This took about 50 min. The second Retention Interval – a week later - involved remembering certain aspects of the stories presented at the first Retention Interval and lasted about 30 min. The last Retention Interval took place a month after the first and again involved retrieval of information presented at the first Retention Interval and lasted about 30 min. At the end of the experiment users were given a small food reward for participating.

Before listening to each story we either gave users a device (CC or PP), or they had to rely on unaided memory (No Prosthesis - NP). In the CC and PP conditions, they obviously also had OM, which they could choose to use instead of the allocated prosthetic device. We instructed them to remember the story either with the assistance of the PM or OM, using whatever memory techniques they would normally use. Users heard each story only once - during the first Retention Interval. We tested memory at 3 different Retention Intervals: same day, 7 days and 30 days later. On each test users had the same prosthesis as when they heard the original story. For instance if a user had access to CC when first listening to the story, we also gave them CC at subsequent Retention Intervals with the same story.

A critical research question was whether and when people made use of devices instead of relying on unaided OM. So, even when users had access to a note-taking device, we made it clear that they were not compelled to use it, and we noted when devices were used in preference to OM.

2.5 Procedure

The experiments were run using a custom built website. Users were first given a general description of the experiment, the stories and the different types of questions that they would be asked as part of each session.

We then gave them a brief web-based, hands-on tutorial providing detailed descriptions of each memory prosthesis and procedures for the experiment. They carried out 3 practice tasks (one with each prosthesis, and one with no prosthesis). The practice tasks were similar to those used in the experiment. Users were allowed to proceed to the actual experiment only if (a) they felt confident with each device and (b) they had successfully completed all practice tasks.

2.5.1 Experimental Tasks

We read users a story with CC, PP or NP depending on the experimental condition. To control for story/Retrieval Method confounds, we counterbalanced the order in which users received stories, the device they used to carry out each task, and the type of question (verbatim/gist) they were asked. Users answered questions on web based forms.

A key research question was the relationship between users' confidence in their organic memory and their note-taking and retrieval strategies. Before answering each memory question we asked users to evaluate their *Confidence* in their ability to answer the question without using the device. The Confidence question was asked *after* the user had read the memory question but *before* they answered it*:* "*How confident are you that you can remember the answer to this question without using your [memory prosthesis name]?"* Responses were generated on 5-point Likert scales.

Users then tried to answer the question. In all conditions we recorded the retrieval time, i.e. how long it took users to answer that question. We also noted whether users relied on OM or the device to answer the question.

Retrieval accuracy was scored in the following way. We first generated an evaluation metric for each question, by having two coders blind to the experimental hypotheses listen to each story twice. They agreed a set of target answers, specifying keywords and context that needed to be present in that answer. Accuracy scores ranged from 0-5 depending on how much of the target answer the user specified. If an answer included all target keywords (or their synonyms) and context, it received a maximum score of 5. Partially correct answers were defined as either (a) containing all keywords, but inaccurate context, or (b) accurate context and incomplete set of keywords. Scoring was carried out independently by the two judges, and disagreements were referred to a third judge for resolution.

2.4.1 Measures and Variables

We collected and report the following data:

- *Accuracy* of answers
- *Efficiency* – how *long* it takes users to recall the relevant information
- *Retrieval Method:* when people had notes available with CC and PP, we looked at whether they relied on their notes or their organic memory to answer a question.
- *Note taking behaviour* - *how* people take notes, Note *Quality* and *Quantity* for CC and PP
- User *Confidence* in their ability to remember unaided.

3. HYPOTHESES AND RESULTS

The results are organised around the following hypotheses:

3.1 Overall Memory Benefits: Do Notes help cue recall?

We compared the NP condition with the two note-taking conditions (i.e. comparing CC, PP and NP scores) to assess overall effects of note-taking devices on Retrieval Accuracy.

We conducted two-way ANOVAs with independent variables 1) *Device*– i.e. which prosthesis was used (PP, CC or none in the NP condition); 2) *Retention Interval* – i.e. length of time since the user heard the story (same day, 7 days later, 30 days later). The dependent variable was *Accuracy*.

Accuracy scores for digital and analogue note-taking devices and NP are shown in Figure 2. As expected there was a significant difference between Devices ($F_{(2,898)}$=78.9, $p<0.0001$ (SD=483.4, mean=241.7)). Planned comparisons showed that there are differences between CC and NP ($p<0.0001$) and between PP and NP ($p<0.0001$) showing the benefits for memory of both types of note-taking prosthesis.

As expected there was a strong interaction between Device and Retention Interval ($F_{(4,898)}$=11.6,$p<0.0001$ (SD=142.1, mean=35.5)), suggesting that the benefit of notes increases after longer periods of time.

We conducted posthoc Tukey analyses at each Retention Interval. We found no significant differences between CC, PP and NP on the first day, suggesting that there were no immediate benefits for notes. But 7 days later, we found a significant difference between both CC and NP ($p<0.0001$) and between PP and NP ($p<0.0001$) – indicating strong short-term benefits of both types of note-taking prostheses. During the final session - 30 days later, we found significant differences between CC and NP ($p<0.0001$) and between CC and PP ($p<0.0001$) – indicating the benefits of digital notes over analogue notes at longer retention intervals, presumably because CC allows access to the underlying verbatim record.

PP and NP were equivalent (p>0.05) showing that the benefits of analogue notes degrades over time.

We conducted further post-hoc tests looking at Accuracy for each Device at the different retention intervals. We found no significant differences for CC for the 3 retention intervals (p > 0.05). But for PP there was a significant difference between 1 and 7 days (p<0.02), and 7 and 30 days (p<0.006). For NP there were differences between 1 and 7 days (p<0.0001) sessions but no differences between 7 and 30 days (p>0.05) sessions. The fact that CC shows no degradation over time, whereas both PP and NP decay shows the benefit of digital note-taking in protecting recall.

Given that notes do indeed help *overall* retrieval our next question was how did so? What are their effects on PM and OM respectively? We first looked at whether note-taking helps *PM* accuracy by generating rich cues that are useful for later PM retrieval.

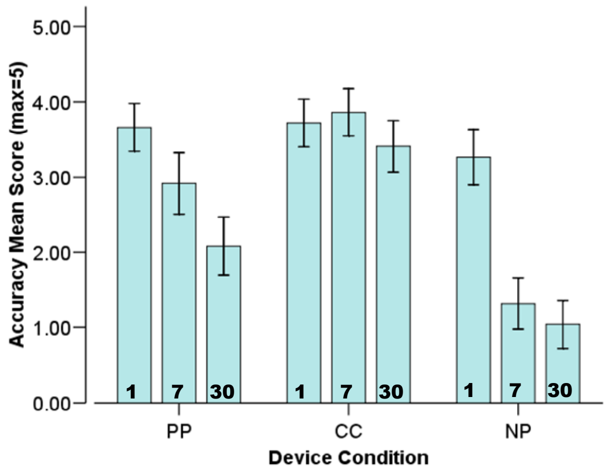

Figure 2. Overall Accuracy for different Device Conditions over 3 Retention Intervals. For each Device the intervals are 1, 7 and 30 days from left to right.

3.2 Do Notes Help Cue PM?

The above analysis does not separate cases for both CC and PP when notes are available but not used from cases where notes are taken and used. It may be that users take notes but choose not to use them for retrieval. To examine this, we compared Accuracy when people actively used CC or PP notes at retrieval with the NP condition when no notes were available. Because we wanted to quantify the direct effects of *notes as retrieval aids*, we excluded from the analysis, cases where people had digital notes, but chose to rely on their unaided memory, as in these cases there was no direct prosthetic use of notes. We conducted an ANOVA with *Retention Interval* (same day, 7 days, 30 days) and *Device* (CC, PP, and NP) as independent variables. *Accuracy* was the dependent variable. Figure 3 illustrates our findings.

As expected, there were significant differences between Devices for Accuracy ($F_{(2,\ 683)}$ = 65.7, p<0.0001 (SD=398.4, mean=199.2)). Planned comparisons of CC and PP, with NP showed the advantage both of digital and paper notes over NP (both p<0.0001).

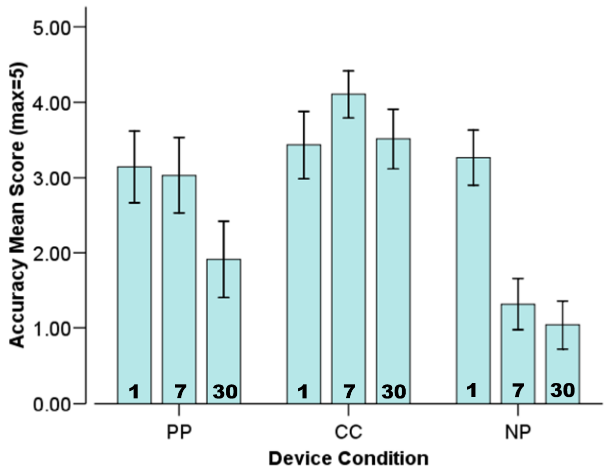

Figure 3. PM Cuing at different Device Conditions over 3 retention Intervals. For each device the intervals are 1, 7 and 30 days from left to right.

There was also an ANOVA interaction between Device and Retention Interval for Accuracy ($F_{(4,683)}$=14.0, p<0.0001 (SD=170.3, mean=42.6)). Post hoc Tukey tests showed that when CC and PP notes were used, performance was better at longer retention intervals. When we compared CC, PP and NP at 1 day, there were no significant differences (all p>0.05). At 7 days, there was a significant difference between CC and NP (p<0.0001) and between PP and NP (p<0.0001). Similarly at 30 days there was a significant difference between both CC and NP (p<0.0001) and between PP and NP (p<0.007). The data show the benefits of actively using notes as PM cues: PP notes, or with CC the combination of notes and access to the verbatim record, help by cueing PM at longer retention intervals, as OM degrades. Further post-hoc tests showed a significant benefit of using CC over PP (p<0.001) notes to access a verbatim recording of a conversation.

We conducted further post-hoc tests examining CC at (1, 7 and 30 days) and found no significant differences (p>0.05). But when we looked at PP, we found a significant difference between 7 and 30 days (p<0.004), although there were no differences between 1 and 7 days (p>0.05). Further post-hot tests with NP revealed significant differences between 1 and 7 day (p<0.0001), but not between 7 and 30 days (p>0.05).

In conclusion, active use of notes do help memory by providing rich cues for PM. In particular, notes combined with a digital record (ie CC) seemed relatively immune to decay. If actively used, PP notes are equally useful immediately and after a week, but their utility degrades at 30 days, even though they are still better than NP overall.

3.3 OM: Do notes distract or help focus OM?

There are competing views about how notes affect organic memory. To investigate whether taking notes *distracts,* or *focuses* OM, we examined the differences in Accuracy when people had taken notes *but chose not to use them* preferring to rely on their unaided memory to remember. We compare unaided OM in the CC and PP conditions (when notes have been taken but not used), with NP scores where no such notes are available. We conducted an ANOVA with *Retention Interval* (same day, 7 days, 30 days) and *Device* (CC, PP, and

NP) as independent variables. *Accuracy* was the dependent variable. Figure 4 illustrates those findings.

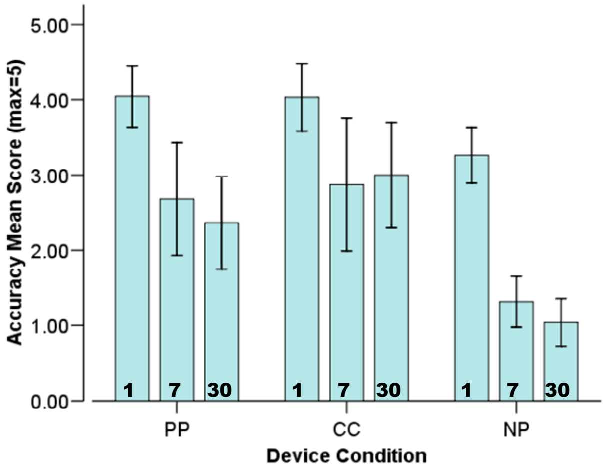

Figure 4. OM focus at different Device Conditions over 3 Retention Intervals. For each device the intervals are 1, 7 and 30 days from left to right.

As expected, there was a significant difference between Devices for Accuracy ($F_{(2, 514)} = 32.1$, $p<0.0001$ (SD=191.1, mean=95.6)) with planned Tukey comparisons showing that in cases where people had taken either CC or PP notes they outperformed NP for unaided memory ($p<0.0001$).

However, there was no interaction between Device type and Retention interval ($F_{(4, 514)} = 1.7$, $p>0.05$ (SD=20.1, mean=5)). As Figure 4 indicates the differences between devices are equivalent at each retention interval. We conducted further post-hoc tests looking at all PM devices at (1, 7 and 30 days). In all cases OM dropped between 1 and 7 days with no subsequent differences between 7 and 30 days.

Overall the results support the focusing hypothesis and contradict the distraction hypothesis. Even when people choose not to use their notes at recall, the initial act of taking notes helps to boost their memory compared with when no notes are taken. However, unlike CC cueing of PM, this boosting effect on OM degrades over time.

3.4 Effects of Note-type: Quality vs. Quantity

We next looked at the effects of different types of notes on retrieval. Is it better to take larger numbers of notes (i.e. high *Quantity*) in order to generate exhaustive PM cues? Or should people employ more concise note-taking strategies that try to focus on more critical information (i.e. high *Quality*)?

We scored the *Quantity* of notes simply by counting the number of words that users recorded. *Quality* was more complex. For each story, we devised a marking scheme consisting of the ideal set of notes that would have to be generated to cover all the topics that we asked users about. This included topic keywords plus contextual information about each topic, and both were required to achieve a perfect notes score, see Figure 5. Two independent judges applied the marking scheme; they gave 5 marks for complete notes which captured both keywords (or their synonyms), plus context. Marks were reduced for incomplete and partial information. Consistency between judges was 90% and disagreements were resolved by discussion.

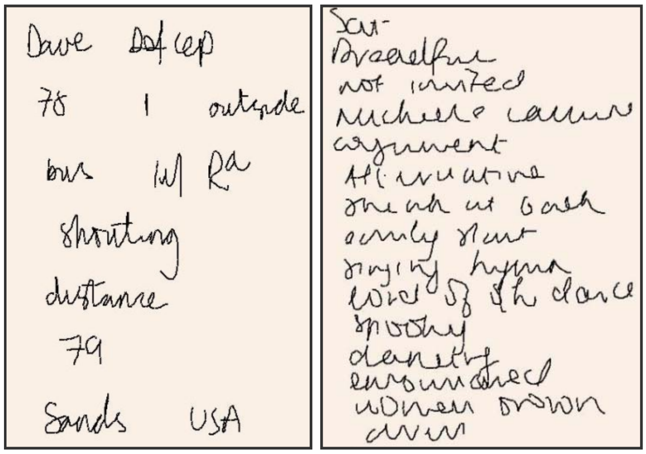

Figure 5. CC low and high quantity notes from two different note takers

Similarities between CC and PP notes

Before examining the effects of note type on memory we looked first at how people took notes digitally compared with pen and paper. For instance, when people used CC, did they take more or fewer notes, in comparison to when they used PP?

We found that people had consistent note-taking strategies - taking similar *Quality* notes digitally and on paper. The strength of this relationship is confirmed by a correlation between *Quality* of digital CC notes and *Quality* of PP notes ($r_{(25)} = 0.4$, $p< 0.001$). Figure 6 shows CC and PP notes from the same note taker) – suggesting similar note-types and strategies in both cases.

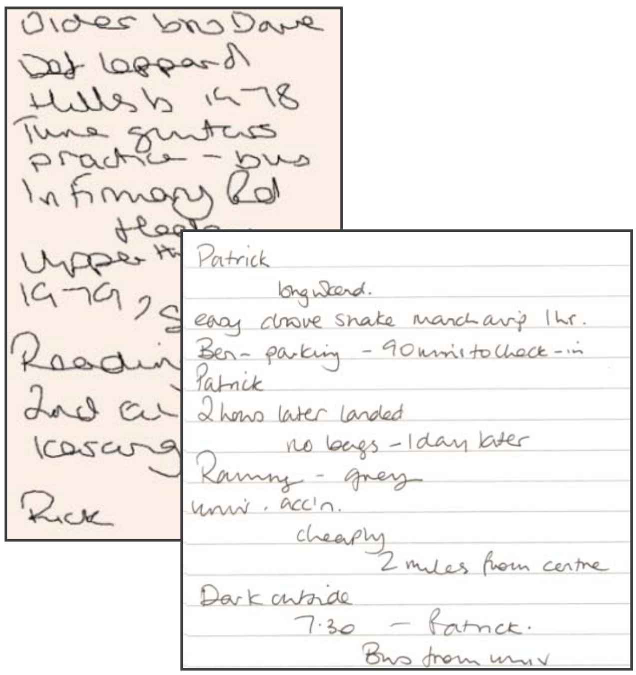

Figure 6. CC and PP notes from the same note taker.

The *Quantity* of notes taken across digital and paper media was also consistent. Again, there was a strong positive correlation between the *Quantity* of CC and PP notes ($r_{(25)} = 0.6$, $p < 0.001$).

We were also interested in how people used the space on the Digital PDA screen compared with the paper sheet. We expected people to spread their notes out more on paper but to have more condensed digital notes. We measured space usage in terms of the number of lines used, expressed as a percentage of the number of available lines.

Contrary to our expectations, we found that people used space in a similar way for digital and paper notes. There was a strong correlation between CC and PP space usage ($r_{(1199)} = 0.31$, $p<0.001$) suggesting that people tend to apply familiar note-taking strategies to new technology.

We also observed that people used often bullet points to arrange their notes in both CC and PP. Again we found a strong correlation between bullet point use in CC and PP ($r_{(25)} = 0.28$, $p<0.001$).

Having established that digital and pen and paper notes are similar we combined digital and analogue notes in testing the relations between note type and memory.

Quality:

We looked at whether people who take high *Quality* notes in both CC and PP also remember better overall. We found a significant correlation between *Quality* of notes and overall memory – regardless of whether people answered memory questions prosthetically using their notes or relying on unaided OM ($r_{(515)} = 0.4$, $p<0.001$).

Quantity:

In contrast Quantity was not such a strong predictor of memory. While taking a higher Quantity of notes overall also improved retrieval, the overall correlation was much weaker ($r_{(515)} = 0.1$, $p<0.05$).

We also tested whether Quality was a better predictor of recall Accuracy than Quantity and found that this was indeed the case. A comparison of the correlation coefficients using the Hotelling/Williams Test showed that Quality was much the stronger predictor ($t_{(515)} = 4.63$, $p < 0.0001$).

Retrieval Efficiency:

We also looked at the effects of note *Quantity* on *Retrieval Efficiency*. Does taking more notes increase speed of retrieval or are too many notes distracting as retrieval cues?

There was a strong positive correlation between overall note *Quantity* and Efficiency (time to answer each question), $r_{(384)} = 0.3$, $p<0.001$. People who took more notes tended to take *longer* to generate responses using PM. This suggests that having a large volume of notes decreases speed of retrieval - as there are more notes to scan to find a promising index.

Reasons for Note-Taking:

Finally we looked at what motivates people to take more notes. Do people who are *less confident* take larger volumes of notes, or does confidence result from having good notes?

Contrary to our hypothesis, there was no significant correlation between combined CC and PP note *Quantity* and overall *Confidence* scores ($r_{(515)} = -0.003$, $p>0.9$). People who are unconfident about their memory don't act upon this information to take more notes.

However, there was a strong positive correlation between overall note *Quality* and *Confidence* scores ($r_{(515)} = 0.3$, $p<0.001$). This may be because having higher *Quality* notes seems to boost people's confidence that they will be able to remember unaided. Or alternatively people who have better memories tend to take better notes and they are more confident about OM based on their past success of remembering.

3.5 Subjective User Comments

Overall people voiced a liking for memory cueing techniques such as note taking. The majority of users preferred digital over paper notes but this could be due to the audio back up provided alongside the digital notes "*[CC] is similar to writing [PP] notes which I like plus back up of recording*". Nevertheless CC was appreciated for its familiarity to PP and its ease of use "*[CC] is very easy to use and it's very accurate*".

But after some time had elapsed people worried that their notes might not be sufficient to guarantee long term retrieval because they were contextually dependent on fallible organic memory to interpret them "*my notes were a mess, and were contextual with my own memory, which itself had faded. Hence the usefulness of the notes was severely undermined*".

A few people acknowledged the importance of taking high Quality notes "*[PP] jogs your own memory. It depends largely on the quality of notes*"…and "*you have notes for prompts if you make the right prompts!*"

Some users discussed the *Quantity* of their notes and the need not to take too many notes "*[after 7 days] the notes I had done were the right amount - not too lengthy*". As we reported earlier, having a large *Quantity* of notes was not perceived as an effective memory cue as it takes too long to retrieve relevant information.

There was also concern about note taking technique "*I've never been very good at making notes*" and "*…the notes are only as good as the user…*" But even with poor note taking skills, people realised that having some notes - whatever their quality - might be better than having no PM backup at all "*I'm dependent on my own ability to make notes but still better than memory alone*".

4. CONCLUSIONS

This study examined whether users' attempts to prepare for future retrieval using a specific type of prosthetic device, namely notes, led to improved recall.

What can our results tell us about digital memory more generally? We have shown that there are two independent mechanisms by which users preparations for future retrieval can have effects on memory:

(1) PM cueing – by generating useful cues (notes) that trigger memories when users access them at retrieval. In this study, digital notes in CC were highly robust as retrieval cues showing minimal decay over the month of the study

(2) OM focusing – the very act of generating cues helps memory (even if these cues are never consulted). However, these focusing benefits decay over time.

Also we found some effects for the *types* of cues that people construct. Higher quality cues helped retrieval whereas large volumes only weakly did so. Furthermore there were *costs* to taking too many notes. Generating too many cues leads to more inefficient retrieval with increased retrieval times when users generated very many notes.

Finally we clarified the relationship between note-taking behaviour and confidence. People who lacked confidence in their OM were no more likely to take large numbers of notes than those who were very confident. Rather it seemed that taking good notes caused people to be more confident that they would remember unaided.

There were also few observed differences between digital and pen and paper note-taking practices. Digital and analogue notetakers tend to exploit space in similar ways, to use equivalent numbers of bullet points and to take similar volume and quality of notes.

These results have important implications for other studies of prosthetic memory. There has been much recent interest in techniques that allow digital memories to be automatically indexed [33]. Now while such automatic techniques may prove useful, our focusing results show that having users generate *their own cues* is helpful rather than distracting, even when these aren't directly used at retrieval. We did not directly contrast automatic versus manual indexing in this study. At the very least, however, our results suggest that in addition to such automatic techniques we need lightweight ways for people to construct their own retrieval cues, because of the demonstrated benefits they bring.

The confidence results suggest that having high quality personal indices increases confidence and hence the likelihood that the digital memory will be used. Finally the demonstrated relationship between Quantity and Efficiency showed that we need to be careful about how many cues that we generate (whether this is done automatically or manually). Too many cues reduce the efficiency of retrieval.

More specifically the study shows the benefits of a new type of digital note-taking device CC for helping memory. It is more robust than both PP notes and unaided memory. In the spirit of Web2.0, we are also currently investigating extensions to the device which will allow collaborative sharing of notes allowing for example a class of students to share digital notes that are indexed to a podcast lecture [7]. Further we are evaluating a different version of CC that uses pictures rather than annotations and looking at how different *types* of annotations such as pictures support retrieval. This work extends recent studies of how pictures help individuals remember events from their everyday lives [25].

In conclusion we have shown the benefits of new types of digital note-taking prostheses in helping memory and clarified some of the different mechanisms by which they achieve their effects. Future work needs to extend these questions to look at how manual cueing contrasts with automatic methods and how well these techniques generalise to other types of indices such as pictures [25] or more complex narratives [10].

5. ACKNOWLEDGMENTS

We thank our participants for their time and in sharing their stories, memories and giving their feedback.

6. REFERENCES

[1] Abrams D., Baecker R., Chignell M., Information archiving with bookmarks: personal Web space construction and organization, Proceedings of the SIGCHI conference on Human factors in computing systems, p.41-48, April 18-23, 1998, Los Angeles, California, USA.

[2] Abowd G. D. , Harvel L.D., Brotherton J.A., Building a Digital Library of Captured Education Experiences. In Proc. ICDL 2000, Kyoto, Japan (2000), 395-402.

[3] Barnett J., Di Vesta F., Rogozinski J. What is learnt in note taking? Journal of Educational Psychology (1981), 23(2), 181-192.

[4] Benton S., Kiewra K., Whitfill J., Dennison R. Encoding and external-storage effects on writing responses. *Journal of Educational Psychology* (1993), 85. 267-280.

[5] Bellotti V., Dalla B., Good N., Bobrow D. G., Ducheneaut N. What a to-do: studies of task management towards the design of a personal task list manager. *ACM Conference on Human Factors in Computing Systems (CHI2004)*; 2004 April 24-29; Vienna; Austria. NY: ACM; 2004; 735-742.

[6] Bellotti V., Ducheneaut N., Howard M., Smith I., Taking email to task: the design and evaluation of a task management centered email tool, Proceedings of the SIGCHI conference on Human factors in computing systems, April 05-10, 2003, Ft. Lauderdale, Florida, USA.

[7] Davis R.C., Landay J.A., Chen V., Huang J., Lee R. B., Li F.C., Lin J., Morrey III C. B, Schleimer B., Price M. N., Schilit B.N. NotePals: lightweight note sharing by the group, for the group, Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, p.338-345, May 15-20, 1999, Pittsburgh, Pennsylvania, USA.

[8] Dickie C., Vertegaal R., Chen D., Fono D., Cheng D., Sohn C. Augmenting and Sharing Memory with EyeBlog. *In Proc. CARPE 2004.* NYC, ACM Press (2004), 105-109.

[9] Dumais S.T., Cutrell E., Cadiz J.J., Jancke G., Sarin R., Robbins D.C. Stuff I've Seen: A system for personal information retrieval and re-use. *In Proc. SIGIR 2003,* Toronto, Canada, ACM Press (2003), 72-79.

[10] Frohlich D.M., Kuchinsky, A., Pering C., Don A. & Ariss S. (2002) Requirements for photoware. *Proceedings of CSCW '02,* New York: ACM Press.

[11] Gemmell J., Bell G., Lueder R. MyLifeBits: a personal database for everything. *In Comm. of the ACM, 49 (1)* (2006), 88-95.

[12] Jones W., Bruce H., Dumais S., Keeping found things found on the web, Proceedings of the tenth international conference on Information and knowledge management, October 05-10, 2001, Atlanta, Georgia, USA.

[13] Kalnikaite, V., and Whittaker, S. (2007). Software or Wetware? Discovering When and Why People Use Digital Prosthetic Memory. In *Proc. CHI07*, ACM Press (2007), 71-80.

[14] Kidd A. The marks are on the knowledge worker. *In Proc. CHI94*, ACM Press (1994), 186-191.

[15] Kiewra K., Mayer R., Christensen M., Kim S., Risch N. Effects of repetition on recall and note-taking: Strategies for learning from lectures. *Journal of Educational Psychology*, (1991) 83. 120-123.

[16] Kirk, D., Sellen, A., Rother, C., and Wood, K. (2006). Understanding "photowork". *Proceedings of CHI 2006.*

[17] Lamming M., Flynn M. "Forget-me-not" intimate computing in support of human memory. *Tech rep. Rank Xerox, EuroPARC*, Cambridge, UK (1994), 94-103.

[18] Lachman M.E., & Leff R., (1989). Perceived control and intellectual functioning in the elderly: a 5 year longitudinal study. Developmental Psychology, 25, 722-8.

[19] Landay J.A., Kaugmann T.R. User Interface Issues in Mobile Computing. *In Proc. WWOS-IV 1993*, IEEE Press (1993), 40-47.

[20] Mann S., Fung J., Aimone C., Sehgal A., Chen, D. Designing EyeTap Digital Eyeglasses for Continuous Lifelong Capture and Sharing of Personal Experiences. *In Proc. CHI 2005 Conference on Computer Human Interaction*, ACM Press (2005).

[21] Malone, T.W. How do people organize their desks? Implications for designing office information systems, *ACM Transactions on Office Information Systems*, 1983, *1*, 99-112

[22] Petrelli D., Whittaker S., and Brockmeier J. AutoTopography: What can Pshysical Mementos Tell Us about Digital Memories? In press to appear in CHI 2008, Conference on Human Computer Interaction, NY: ACM Press.

[23] Piolat A., Olive T., Kellog R. Cognitive effect during note taking. Applied Cognitive Psychology (2005), 19:3, 291.

[24] Rodden K., Wood K. R., How do people manage their digital photographs?, Proceedings of the SIGCHI conference on Human factors in computing systems, April 05-10, 2003, Ft. Lauderdale, Florida, USA.

[25] Sellen A. J., Fogg A., Hodges S., Wood K. Do Life-Logging Technologies Support Memory for the Past? An Experimental Study Using SenseCam. *In Proceedings of CHI 2007.*

[26] Sellen, A., Fogg, A., Hodges, S., Rother, C. and Wood, K. (2007). Do life-logging technologies support memory for the past? An experimental study using SenseCam. *Proceedings of CHI '07.* Honourable Mention, Best of CHI Awards.

[27] Stifelman L., Arons B., and Schmandt C. Augmented real-world objects: a paper-based audio notebook. In Proc. CHI 1996 Conference on Computer Human Interaction, Vancouver, Canada, 2001, ACM Press, New York, NY (2001), 182-189.

[28] Tauscher and L., Greenberg S. How people revisit web pages: Empirical _ndings and implications for the design of history systems. International Journal of Human Computer Studies, Special issue on World Wide Web Usability, 47:97, 138, 1997.

[29] Whittaker S., Hyland P., Wiley M. Filochat: handwritten notes provide access to recorded conversations. In Proc. CHI 1994, Conference on Computer Human Interaction, Boston MA. ACM Press, New York, NY (1994), 271-277.

[30] Whittaker S., Hirschberg J. Look or Listen: Discovering Effective Techniques for Accessing Speech Dat*a. In Proc. CHI98*. Springer-VErlag (2003), 253-269.

[31] Whittaker S., Sidner C. Email overload: exploring personal information management of email. *In Proc. of CHI 1996 Conference on Computer Human Interaction*, NY: ACM Press, 276-283.

[32] Whittaker, S., Swanson, G., Kucan, J., & Sidner, C., (1997). Telenotes: managing lightweight interactions in the desktop. ACM *Transactions on Computer Human Interaction*, 4, 137-168.

[33] Wenyin, L., Dumais, S.T., Sun, Y.F., Zhang, H.J., Czerwinski, M.P., Field, B. Semi-automatic image annotation. In Proceedings of Interact 2001, Eighth IFIP TC.13 Conference on Human Computer Interaction. July 2001.