Learning From Our Mistakes – Identifying Opportunities for Technology Intervention to Address Everyday Cognitive Failure

Learning From Our Mistakes – Identifying Opportunities for Technology Intervention to Address Everyday Cognitive Failure

Sarah Clinch and Cecilia Mascolo

Abstract—There is growing opportunity for technologies to augment human memory and other cognitive processes – widespread pervasive sensing enables fine-grained traces of human activity, and advances in display technologies enable systems to provide input to user cognition in a wide array of settings. However, systems to date typically either address known cognitive impairment (e.g. autism, Alzheimer's disease), or look to enhance ones' general capacity for a specific task. In contrast to these approaches, we argue that *a focus on recognition and quantification of human error* is key to the design of future systems for augmenting the human mind. By focusing on the errors made in cognition, we can first identify frequent, persistent or severe failures as targets for future pervasive computing systems, and then go on to measure the success of any interventions developed, i.e. by asking *have the augmentation systems delivered actually reduced the prevalence, persistence or impact of a specific cognitive error?*

In this article, we make the general case for the study of human error in order to support the design and evaluation of technology interventions intended to extend cognitive capabilities, before focusing on a case study in the augmentation of human memory.

———————————————————————

1. Introduction

Early toolmaking and industrial innovation sought to extend the physical capacity of man. By contrast, the emergence of computers has clear potential to extend not just the physical capacity of humans, but the capacity of the human mind [17].

Human cognition encompasses a range of processes from managing attention, to acquiring knowledge, problem solving, storing and retrieving memories, evaluating risk, and making decisions (amongst others). Whilst digital computation may surpass and replace some human cognition (e.g. the replacement of mental arithmetic with use of an electronic calculator), a more compelling challenge is arguably the realisation of 'tools' that seamlessly extend cognitive capabilities, giving the illusion (for example) of a 10% increase in IQ, human memory or decision making capability. Widespread pervasive sensing already enables fine-grained traces of human activity; likewise advances in display technologies (e.g. pervasive displays, augmented reality) allows technology input to user cognition in a wide array of settings.

As an example, recent research attention has been given to the augmentation of human memory [6]. Building on developments in life-logging, quantified self, and personal information presentation (e.g. smartphones, watches, heads-up displays), a common approach sees wearable cameras and other devices capture a rich feed representing the human experience that can then be summarised and presented back to the user to help cue retrieval and strengthen connections between the neurons that form a memory trace.

Delivering value in augmented human mind requires researchers and developers to understand and address two issues. Firstly, to identify the role of technology in supporting cognition, establishing the points at which technical interventions could deliver value to users by extending their natural capacity and addressing known limitations or vulnerabilities in their thinking. Secondly, to monitor the success of developed systems, verifying their effectiveness in extending mental capabilities both in lab settings and in real world lived experience. To date, the first of these has been largely technology-driven, whilst the second has predominantly relied on lab-based testing (e.g. [10, 11]).

In this article, we propose that by focusing on "everyday cognitive failures" [2], researchers can both identify key opportunities for technical intervention and establish on-going measurement, allowing them to iterate developed systems and deliver tangible improvement in uncontrolled daily-living settings. Our article makes the general case for the use of human error in technology intended to extend cognitive capabilities, before focusing on a case study in human memory in which a diary study capturing everyday memory failures is used as an illustration of how cataloguing errors in mental processes might suggest, and shape the design of, future augmentation systems.

2. Utilising Failure

The human-computer interaction community has long recognised the importance of cognition in the design of computer systems. Users build mental models of technologies and software that, in turn, shape their interactions

with those systems – in cases where a platform performs differently to the user's understanding, their actions may result in errors. This then, was what led Donald Norman to articulate his case for researchers, developers and designers to:

> *"Use analyses of people's performance in a variety of situations–but especially their errors–to construct an analysis of the appropriate form of human-machine interface that would optimize performance and minimize either the incidence of error or the effect of the error"*.

> – *Donald Norman [12]*

Although Norman's focus was on the study of human errors when engaging with machines, here we make the case that understanding errors in everyday cognition and action, separate from interaction with any specific technology, also offers potential for informing the design of future systems aimed specifically to support the capabilities of the human mind. Specifically, we argue that monitoring everyday cognitive failures *(i) informs design*, by suggesting new opportunities for technology interventions that expand the capabilities of the human mind, and *(ii) closes the loop*, providing on-going feedback to evaluate the success of developed interventions and further improve their utility.

2.1 Informing Design

Designing systems to intervene in cognitive processes and extend the capabilities of the human mind raises considerable new challenges. Despite two decades of experimentation with technologies for cognitive intervention (e.g. [8, 15, 16]), the form of successful systems is still unknown. This largely stems from the fact that, despite continuous development in our understanding of specific cognitive impairments (e.g. autism, Alzheimer's), we are yet to develop a full understanding of healthy cognitive processes and how they may be emulated and enhanced.

As a key mechanism for developing this understanding, we identify a need for research that focuses explicitly on the cognitive limitations of both the healthy and impaired mind. Studying both the cognitive limitations, and the human experience associated with reaching one's limits, offers new potential to inform the design of systems intended to overcome them. Such understanding could, for example, answer questions such as:

- When should technology intervene in cognition?
- What technology mediums are best suited to cognitive intervention?
- What cognitive processes should technology facilitate?
- How should social context be incorporated into the design of cognitive interventions?

Rather than simply looking to apply existing technologies to aspects of cognition, an exploration of failures in thought might lead researchers to identify key requirements for new cognitive technologies, driving innovation in both hardware and software.

2.2 Closing The Loop

Measuring the success of cognitive interventions is extremely challenging, and current approaches typically suffer from one or more of the following limitations:

- **Dependence on pre-defined stimuli or inventories rather than naturally occurring cognitive processes**. Psychologists and clinicians use inventories and tightly defined laboratory tasks for measuring cognition (e.g. Broadbent *et al*.'s Cognitive Failures Questionnaire [2]). Computer scientists typically take similar approaches when evaluating systems for human cognition, using well-defined measures and comparing performance with and without intervention. Whilst these provide a mechanism for evaluation, they are often somewhat arbitrary with little or no attempt to map them to real-world utility. For example, our own research [11] establishes the Recall Performance Measure (RPS_d) to evaluate improvements in episodic memory for daily activities following review with egocentric video summarisations. RPS_d aggregates scores representing the richness of seven aspects of recall across multiple events occurring in a time period. The focus here is purely on richness of description, rather the proportion of experienced events recalled or any sense that the participant has increased capability to remember the events that they are most keen to preserve.
- **Reliance on Specialist Equipment**. As an alternative to behavioural measures, computer scientists have explored the use of physical measures of cognitive activities captured through neuro-physiological sensors that directly monitor cerebral activity (e.g. electroencephalography) or through indirect measures such as eye-movement, heart rate and galvanic skin response. These sensing approaches are commonly seen in Augmented Cognition systems that leverage real-time cognitive assessment to provide feedback

to systems used in settings where users are at high risk of cognitive overload [8]. However, biophysical data typically requires considerable expertise to interpret and is difficult to map to internal state. Moreover, the most direct measures of cognitive activity (i.e. neuro-physiological sensors) are costly and tie users to pre-determined instrumented settings.

- **Reliance on Short-term, Laboratory-Based Evaluation**. The above two limitations, together with the challenges of longitudinal in-the-wild research [3] lead many researchers to conduct short-term and/or laboratory-based evaluative studies. Compounding this issue, is the immaturity of research in the augmentation of human cognition – developed interventions often take the form of small-scale prototypes that may not be able to support longer-term and larger-scale studies in uncontrolled settings.

As an alternative to relying on artificial stimuli, we propose a focus on naturally occurring errors, ensuring that the evaluation of developed systems considers improvements seen in daily cognitive activities. Our case study demonstrates that these can easily be measured outside of the lab, with no specialist equipment. Although our case study is relatively short in duration (26 days), and relies on analogue capture, the use of smartphone-based mobile experience sampling tools [13] could easily provide the means to engage participants over extended time periods. Furthermore, by enabling longer-term in-situ studies, a focus on cognitive failures also provides a means by which to close the loop in deployable platforms for cognition. As researchers develop more mature systems that can be deployed in the wild, on-going measurement of observed cognitive failures can provide feedback that determines the degree to which an intervention is improving cognition. This data will enable researchers to engage in iterative design and development, as well as introducing the potential for machine learning and similar techniques to produce cognitive interventions that learn from their successes and failures.

Cognitive errors are varied in nature and severity, affecting any and all of our mental processes (decision making, motor control, problem solving etc.). To better understand how studying human errors can inform the design and evaluation of future systems, the remaining sections move from general cognition to a specific study of human memory, and how examples of how identified failures could suggest new directions for design in memory augmentation. Whilst the study cannot answer all of the questions raised in Section 2.1, it serves as a valuable illustration of how capturing data about everyday experiences of failed cognition may move researchers closer to developing systems that address tangible deficits in memory.

3. A Case Study in Everyday Memory Failure

Psychologists describe human memory as a combination of systems. *Retrospective memory* concerns our learning from past experiences including *semantic* (memory for facts learned over time), *episodic* (autobiographical), and *procedural* (motor skills). By contrast, *prospective memory* relates to memory for future plans or intentions. Each of these memory systems has been shown to be vulnerable to failures – incidences commonly referred to as moments of *everyday forgetting* [14, 20].

Increased problems with forgetting are associated with the cognitive decline of ageing, with trauma, and with brain disease. However, moments of everyday forgetting are common in both healthy and clinical populations. For example, most readers will undoubtedly identify with scenarios such as:

- entering a room and immediately being unable to recall the purpose for which you came into the room,
- forgetting a person's name when referring to them in a conversation with another person, or
- temporarily forgetting how to complete a previously familiar task.

Undoubtedly, systems for human memory augmentation could have a significant impact for users if they not only gave the illusion of a richer, extended memory, but also directly addressed some of these everyday memory failures. Indeed, by taking cues from the most common, and most severe, incidents of forgetting, system designers can direct their efforts in developing augmented cognition platforms that directly address tangible limitations of human cognition and thus deliver immediate benefit.

Within psychological and medical contexts, techniques to understand the incidence, types, and impact of memory failure typically rely on questionnaires that ask the respondent to reflect back on specific types of failures and the frequency with which they occurred in a specified time period (e.g. *In the last six months, how often did you find you couldn't quite remember something although it was "on the tip of your tongue"?*). Such approaches suffer from significant flaws (see sidebar); as an alternative, a small number of diary studies have provided a larger corpus of everyday memory failure incidents, but the most popular of these is now out-dated, and more recent studies have tended to focus on older populations. There is therefore little up-to-date evidence identifying everyday memory failures in healthy populations that might be the target of technology-based augmentation of cognition.

As a case study in the use of cognitive failure to inform the design of future memory augmentation systems, we conducted a three-week diary study with 14 participants to elicit specific instances of *everyday memory failure[1]*. We recruited a gender-balanced sample of healthy volunteers (< 60 years of age) that included academic, administrative and research staff from four universities (n=8); research students (n=4) from two universities; and other adults in full-time employment (n=2). Participants were recruited on an on-going basis and participated for up to three weeks.

Our study procedure is updated variant of early diary studies used to capture memory failures in psychology (and as such suffers from the usual limitations of self-report studies). Participants were sent emails daily for up to 26 days, and asked to make a note each day of "all occasions on which you notice that you have forgotten something today". Participants were asked to include the time and location of event, a description of the context and the degree to which they found the forgetting incident to be problematic (on a scale from 1-10 where 1 indicates that the incident "didn't really impact you at all", and 10 that it "required substantial time and effort to deal with"). They were provided with three illustrative examples that included all requested details (forgetting to join a teleconference, failing to recall the correct word in a second language, and leaving an item behind).

In addition to reporting memory failure incidents, participants were asked to use free text to summarise the extent to which they thought they had captured all the things they had forgotten that day, and to describe the degree of mental and physical challenge over the day. Finally, participants were also asked to provide five words that best described their mood and energy levels for the day.

Our final dataset consisted of 164 emails reporting a total of 184 study participation days (some participants batched their responses for several days into a single email). Just under half the participants (6) sent 5 or fewer emails (mean 3.00, σ=1.63) that reported a combined total of 7 or fewer days (mean 4.17, σ=1.86); the remaining 8 sent 10–26 emails (mean 18.25, σ=5.24) that included reports for 14 or more days (mean 19.88, σ=4.04).

Sidebar: Everyday Forgetting

Unlike prior research into unusual forgetting phenomena emerging as a result of trauma or brain disease, recent decades have seen psychologists look with growing interest at "normal, everyday remembering and forgetting" [20]. Studies of everyday forgetting typically adopt one of two methods (1) subjective questionnaires that reflect back on memory and forgetting over time, and (2) diary studies that catalogue moments of forgetting on a daily basis.

Retrospective Questionnaires

Parallel work by Herrmann and Neisser [9], and Bennett- Levy and Powell [1] led to development of the Inventory of Memory Experiences (IME) and Subjective Memory Questionnaire (SMQ) respectively. Both questionnaires feature ~50 items of the form "How good is your memory for…" or "How often do you forget…" answered using a rating scale. Broadbent *et al*.'s Cognitive Failures Questionnaire [2] covered broader measures of everyday failures in cognition, incorporating not only failures of memory, but also those relating to attention and physical coordination. This widely adopted tool asks individuals to reflect back over six months, considering how often specific failures occurred (rated on a scale from *never* to *very often*). Finally, the Prospective and Retrospective Memory Questionnaire [4] quantifies incidences of both prospective and retrospective memory over short and long terms, with or without relevant cues.

Although valuable, these inventories have frequently been shown to demonstrate a number of issues, delivering paradoxical results (e.g. respondents frequently reporting "above average" memory [1], and cognition that improves with age [7]).

Diary Studies

Studies have attempted to address the problem of retrospective, self-evaluative reporting by providing participants with the means to capture moments of forgetting in an ongoing record (typically a paper diary). These diary studies operate over a shorter timescale (e.g. two weeks) but allow richer capture in the form of one or more sentences per incident.

Two such studies were conducted by Crovitz and Daniel [5], and Terry [20], both engaging around fifty young adults and generating up to 1000 moments of forgetting for further study. Crovitz and Daniel reduced their participants' diary entries down to 33 synopses that accounted for just under half (492) of reported the moments of forgetting – the most frequently occurring of these was "I forgot a person's name" (semantic, n=113), followed by "I forgot to make a phone call" (prospective, n=73) and "I forgot a phone number" (semantic, n=30); overall about half of the diary entries related prospective failures. Likewise Terry [20] coded entries into 36 broad categories, with the majority being failures of prospective memory.

---

[1] The study went through an institutional ethical approval prior to recruitment; participation was rewarded with entry into a draw to win one of two online shopping vouchers (value ~$60).

More recently, Ramos *et al.* [14] conducted a two-week study with two cohorts of older adults. 18 participants reported a total of 129 incidents (0-19 per person) with an additional 25 incidents elicited in subsequent discussion. Prospective memory failures were the most prevalent (40%), followed by semantic (22%) episodic (22%), attention (16%) and procedural (1%).

The study described in Section 2 adapts the methodology used in the above diary studies, providing an updated corpus of memory failures occurring in healthy adults (accounting for example, for recent changes in technology such as the increased prevalence of email, and a reduced tendency to memorise telephone numbers).

3.1 Number and Frequency of Forgetting Incidents

A significant minority (14.63%) of daily reports contained zero identified incidences of forgetting, and whilst these reports were dominated by two participants, nine participants submitted such a report on at least one occasion, and five more than once. The remaining reports included up to ten moments of forgetting per day (mean 2.78, σ=1.84). Frequency of forgetting incidents varied, but only two participants (one male, one female) consistently reported $\geq$ 5 forgotten incidents in a day (accounting for 27–38% of their reports).

A total of 447 incidents were reported, with the majority occurring between 8am and 11:59pm [Figure 1b]. We see no meaningful difference in hourly reporting patterns for males and females, and little change (a slight reduction) in overall hourly reporting patterns throughout the day. Slightly more incidents are reported mid-week (Tue-Fri) than over the weekend and on Mondays [Figure 1a].

3.2 Type and Severity of Forgetting Incidents

One incident report simply stated that a memory failure occurred, with insufficient detail to categorise further: *"I did forget something, but I've forgotten what it was! It wasn't anything big." [Participant N]*. The remaining 446 incidents were classified according to the type of memory failure that had occurred; of these, two descriptions were too ambiguous to allow classification. Two described an inability to recall the current time of day, and 20 were considered to be failures of attention rather than of memory itself[2]; 9 were episodic failures, 6 procedural, 325 prospective, and 82 semantic.

411 reports included a score describing the degree to which the participant considered their memory failure to have been problematic; a summary of these is given in Table 1. The majority of incidents were not considered problematic, scoring a mean of 2.56 (σ=1.81) out of ten, with 79% of scored incidents being given a value of three or below. Only seven incidents (2%) were scored at an eight or above.

Procedural memory failures were typically considered to be the most problematic (n=6, mean score 3.67, σ=2.05) and *episodic* the least problematic (n=8, mean score 1.75, σ=0.43). Of the more commonly occurring incident types, *prospective* memory errors were generally considered slightly more problematic than average (n=297, mean score 2.70, σ=1.91), and *semantic* slightly less problematic than average (n=79, mean score 2.03, σ=1.22.
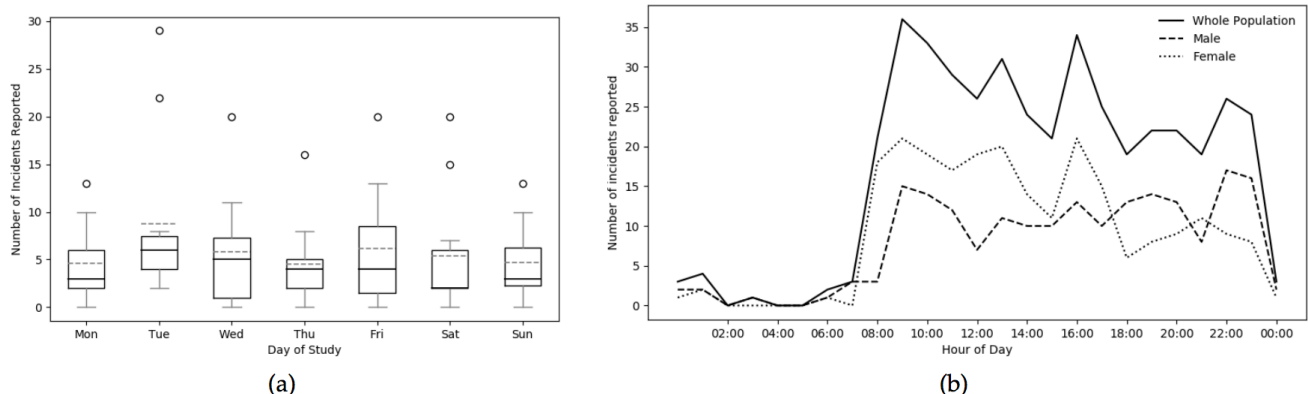

FIGURE 1: Incidents reported over the course of the study: (a) boxplot showing the number of incidents reported by day of week (solid box lines represent the median and quartiles, dashed lines represent the mean, whiskers placed at 1.5 times the IQR); (b) total number of incidents reported by hour of day, with additional breakdown by gender.

---

[2] Although our study explicitly asked participants what it was they had forgotten, like Ramos *et al.* [14] we find a small minority of responses relate more to a failure of attention than of memory. To clarify, whilst forgetting to take one's wallet when leaving the house would typically be considered a failure of prospective memory, forgetting where the wallet placed is often a result of inattention.

Score	attention		episodic		procedural		prospective		semantic		other		Total	
1	5	31%	2	25%	1	17%	96	32%	33	42%	3	60%	140	34%
2	4	25%	6	75%	2	33%	85	29%	26	33%	0	-	123	30%
3	4	25%	0	-	0	-	45	15%	11	14%	1	20%	61	15%
4	1	6%	0	-	0	-	15	5%	6	8%	0	-	22	5%
5	1	6%	0	-	1	17%	24	8%	1	1%	0	-	27	7%
6	0	-	0	-	2	33%	19	6%	1	1%	0	-	22	5%
7	1	6%	0	-	0	-	6	2%	1	1%	1	20%	9	2%
8	0	-	0	-	0	-	1	0%	0	-	0	-	1	< 1%
9	0	-	0	-	0	-	4	1%	0	-	0	-	4	1%
10	0	-	0	-	0	-	2	1%	0	-	0	-	2	< 1%
Total	16	100%	8	100%	6	100%	297	100%	79	100%	5	100%	411	100%
Unscored	4		1		0		28		3		0		36	

TABLE 1: Participant reports indicating the degree to which they *"found the [reported] forgetting incident to be problematic. . . at the time it occurred (on a scale from 1-10 where 1 indicates that the forgetting didn't really impact [them] at all, and 10 that the forgetting required substantial time and effort to deal with)"*.
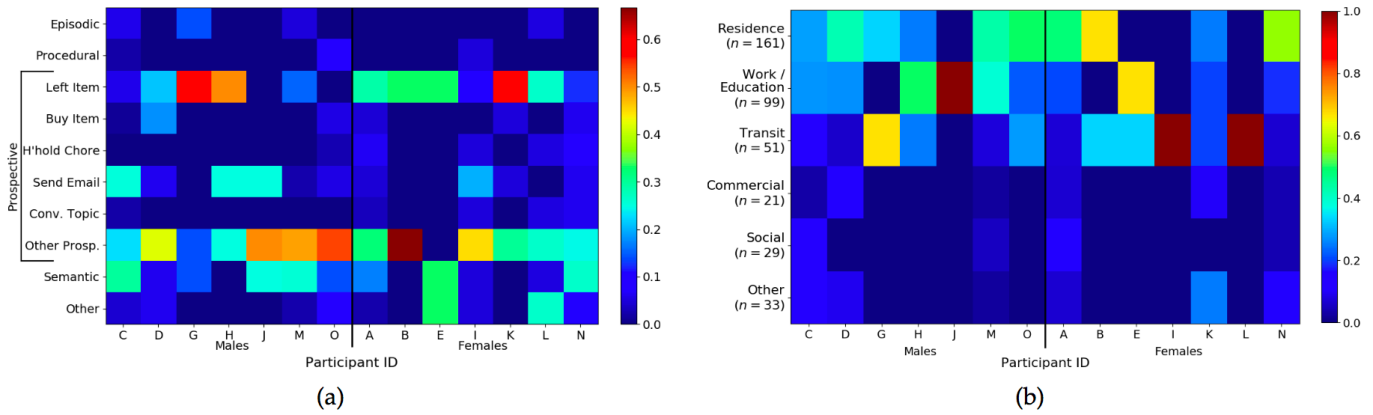

(a)

(b)

FIGURE 2: Heatmaps representing the type and location of memory failure incidents broken down by participant: (a) the proportion of errors of each type reported by participants; (b) proportions of errors reported at each location. Both plots demonstrate overall trends (e.g. a propensity towards prospective errors, particularly those involving left items, and a tendency to report errors at home, work and in transit), but individual differences are also seen.

4. Implications for Design

The described study provides a corpus of memory-related cognitive failures – how then does the resulting dataset shape future design for human memory augmentation systems? In this section we identify three potential implications for the design of human memory augmentations that arise from our results: firstly, a reduced emphasis on episodic memory, secondly consideration for the role of transit and changing context, and finally, opportunities for memory interventions that address individual differences.

4.1 Reduced Emphasis on Episodic Memory

To date, human memory augmentation systems have largely been technology driven, for example using experience capture devices such as the SenseCam and Narrative Clip to support episodic memory. By replaying images captured during daily activities, researchers can provide cues and support users in the rehearsal of episodic memories in order to help compensate for specific defects [10] or simply enable richer recall [11]. However, our dataset suggests that episodic memory should not always be the focus of technology intervention, since these were neither the most frequent failures, nor those considered the most problematic.

Targeting Frequently Occurring Memory Failures
The frequency of reported memory failure types has previously been shown to vary with age, with younger participants reporting problems with prospective memory to be the most frequent, and older participants being more likely to report retrospective memory failures [4]. Within our own dataset, prospective memory failures were by far the most frequently reported incidents (72% of our dataset), with items left behind accounting for 19% of all failures. Such a problem could easily become the target of current camera-based memory augmentation architectures [10, 11], identifying common forgotten objects (in this sample, drinks bottles, keys, pens and prepared meals), many of which have a clearly defined opportunity of forgetting in one's daily routine,

and delivering a targeted reminder. Other common prospective failures included forgetting to send emails (10% of all failures), buy items (4%) or complete other domestic chores (3%), and forgetting a planned topic of conversation (2%). Like left items, many of these failures appear to be plausible targets for contextual reminders – in the case of forgotten emails, we saw numerous repeated failures expressed with growing levels of frustration, simply noting the forgotten mails and triggering a reminder next time the user opened their email client could have been sufficient to mitigate these (although it's worth noting that some forgotten tasks may not be truly issues of recall, but may simply be avoidances of an undesirable task). Contextual reminders were a focus of early memory augmentation systems (e.g. the wearable Remembrance Agent [15]), but with the emergence of new technologies researchers have largely shifted focus to other aspects of memory. However, newer (often vision-based) technologies could address a broad range of prospective failures – not just "remind me to buy milk when I'm at the supermarket", but also "let me know when I'm about to leave my lunch behind".

Failures of semantic memory were the second most frequently occurring – here, the most frequently reported incident was a failure to bring to mind someone's name (17% of the reported semantic failures, and 3% of all incidents overall). Although many failures related to an inability to identify colleagues, we also saw several instances of the so-called 'butcher on the bus' phenomenon in which someone known in one context could not be identified when seen in an alternative context (e.g. encountering a former colleague). Addressing this form of semantic error with technical intervention is not beyond reach – one could easily imagine camera-based systems building a database of reoccurring faces, and either explicitly attempting to identify the individual (e.g. through manual labelling, or querying of social network data), or simply providing a selection of image cues that represent that individual in the settings in which they were previously encountered.

Targeting Problematic Memory Failures

Whilst prospective and semantic failures were the most frequent, participant scoring indicates that procedural memory failures are considered the most problematic, suggesting that this is where technology interventions would be most useful. Recent research has demonstrated the potential for haptic feedback to support passive learning of motor skills [18] and our dataset indicates that this is potentially an important area for technology intervention. Procedural memory failures appear more common for new skills (three such incidents were reported by a single individual who was learning to drive during the study) and for familiar activities that have not been performed in some time. We note however, that our sample includes few procedural memory failures; a larger dataset is therefore needed to indicate if the observed severity of procedural failures is simply a product of the few instances captured in this study being relatively problematic. In the rest of this section, we therefore consider the dataset as a whole (i.e. the most problematic incidents across all memory types).

Our dataset contains sixteen incidents scored at a seven or above. Considering these incidents in detail, we note that those with the following characteristics are most likely to be considered highly problematic:

– Incidents that directly lead to the expenditure of significant time or effort (e.g. having to walk home instead of cycling due to a forgotten bicycle pump).
– Incidents that pose a high risk of financial loss (e.g. leaving the house unlocked).
– Incidents that cause inconvenience to others, cause one to appear inconsiderate or could be perceived as "letting someone down" (e.g. a missed doctor's appointment).
– Repeated incidents, particularly those that may become more pressing, and/or other incidents with an associated deadline (e.g. responding to someone's email).
– Incidents that cause personal embarrassment.
– Incidents that lead to other stressful situations.

Whilst these criteria may not directly suggest technology interventions, understanding the factors that make failures of cognition more or less likely to have a significant negative impact is of clear value to system designers and developers. Predicting the degree to which a potential cognitive failure may be considered problematic could, for example, help to determine how best to leverage limited opportunities for intervention, or when an otherwise inconvenient interruption may be justified.

4.2 Transit

The role of contextual changes in forgetting is a frequently observed psychological phenomenon. Specifically, a movement from one location to another has often been associated with a failure to remember information that was previously able to be brought to mind in the prior location, so-called context-dependent forgetting [19]. This is perhaps no better exemplified than by the common phenomenon of walking to a location to complete a known and specific task, only to find that when entering the location one can no longer remember why one is there:

> *"Forgot why I returned to the kitchen when setting the table. Wanted to bring something out but forgot what it was. Went back out to [. . .] remind myself. Turns out I forgot spoons."*
>
> – *Participant N*

Our own corpus of forgetting incidents demonstrate that a high frequency of forgetting events occur while in transit (i.e. those tagged with locations such as 'walking', 'leaving', in the car/taxi/bus/train/subway, or 'on the way home'). These labels account for 13% of incidents with location information, the third most frequently occurring location (after home and work) [Figure 2b]. On further examination, it seems that these incidents clearly divide into two categories:

1) moments of forgetting that occur during or immediately after a transition (as in the above quote from Participant N), and

2) moments of recall that spontaneously occur during or immediately after a transition, but that actually represent prior moments of forgetting.

Whilst this is partly a general pattern in which the reported locations contained a mixture of those at which the forgetting itself occurred, and those at which realisation of an earlier memory failure occurred, we note that a large proportion of spontaneous recollections of forgotten tasks occur whilst in transit:

> *"Driving to kids birthday party - forgot the present! We were nearly at the party and on time, so going back for it was an 8 mile round trip..."*
>
> – *Participant G*

These two distinct patterns both suggest that developers of memory augmentation systems should give special consideration to moments of human mobility. With regards to data capture, forgetting associated with contextual changes might lead developers to focus on predicting when an individual is about to move from one context to another in order to increase the granularity of recording, and to anticipate potential failures with targeted interventions, whilst the phenomenon of spontaneous recall suggests that lightweight provision for mobile capture of spontaneously recalled items for later action is potentially a straight-forward but valuable contribution. However, the prevalence of cognition failures during mobility also suggests direction for the modalities and mediums used for memory interventions themselves – for example, heads-up displays or those integrated into doorways and places of transition may be particularly valuable.

4.3 Personalised Memory Interventions

To date, human memory augmentation systems have at best distinguished two populations: (i) those with impaired cognition impacted by ageing, disease or trauma, and (ii) those whose cognition is considered to be in the neurotypical (normal) range. Our data, captured from younger and middle-aged adult participants, suggests that there may be individual differences even within so-called "healthy" populations. Two examples of these differences can be seen in Figure 2, in which we look specifically at the differing frequency at which each of our participants report specific types of memory failures [Figure 2a], and the locations at which they report their errors [Figure 2b].

Figure 2a shows that for the majority of participants, leaving an item behind is one of the most common instances of memory failure documented; participants G, H and K are particularly prone to this kind of forgetting. However, some participants rarely or never report failures of this type (e.g. participants C, J and O). Likewise, only a subset of participants report a high proportion of semantic errors (participants C, J, M, E and N), and just a very small number of participants report forgotten emails as a significant proportion of their memory failures (participants C, H and J).

Figure 2b indicates that there may also be differences in the contexts in which individuals are more likely to experience memory failure. Overall, locations are dominated by those in which participants spend the most time (home, work), and then in the previously mentioned transit settings. However, we note that whilst some participants reported a greater proportion of their memory failures at home than at work (participants D, O, A, B and N), for others the reverse is true (participants H, J, M and E); three participants show a disproportionately high number of memory failures reported in transit (participants G, I and L). Further research is needed to determine if this difference is a product of differing portions of time spent in these locations, or of individual cognitive differences. For example, in the same way that some individuals are prone to errors in spatial reasoning, some in logic, and some at creativity etc., so too might some individuals be particularly prone to errors in semantic memory whilst others struggle with accurate episodic recall.

Our study clearly highlights the diversity of memory failures occurring in healthy populations suggesting that as well as addressing specific memory impairments, researchers should design for the individual even in systems designed to extend the capabilities of the wider population.

5. Conclusion

In this article, we have made a case for the closer examination of everyday human errors in the design and evaluation of technology interventions designed to extend cognitive capabilities. Although this paper focuses largely on the exploration of errors in the specific case of human memory augmentation, we believe that a more general consideration of cognitive failures has the potential to impact a wide range of applications designed to extend the capabilities of the human mind – examples include decision making and evaluation of risk, managing attention, acquiring knowledge, creative thinking and general problem solving.

Our argument identifies two important roles for the study of everyday cognitive failures in the development of future systems to support the capabilities of the human mind. Firstly, in informing design, suggesting new opportunities for technology interventions that are grounded in documented failures that are either frequent or impactful in everyday living. Secondly, providing a unique means for closing the loop, with on-going evaluation enabling systems that learn from their successes and failures in altering cognition. To support this argument, an analogue diary study is used a case study of how analysis of cognitive failures can suggest new avenues for the application of technology to cognition, whilst future work using techniques such as mobile experience sampling [13] could easily provide the means for continuous feedback to close the loop.

Acknowledgements

References

[1] J. Bennett-Levy and G. E. Powell. The subjective memory questionnaire (SMQ). an investigation into the self-reporting of 'real-life' memory skills. *British Journal of Clinical Psychology*, 19(2):177–188, 1980.

[2] D. E. Broadbent, P. F. Cooper, P. FitzGerald, and K. R. Parkes. The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, 21(1):1–16, 1982.

[3] B. Brown, S. Reeves, and S. Sherwood. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1657–1666. ACM, 2011.

[4] J. Crawford, G. Smith, E. Maylor, S. Della Sala, and R. Logie. The prospective and retrospective memory questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample. *Memory*, 11(3):261-275, 2003.

[5] H. F. Crovitz and W. F. Daniel. Measurements of everyday memory: Toward the prevention of forgetting. *Bulletin of the Psychonomic Society*, 22(5):413–414, 1984.

[6] N. Davies, A. Friday, S. Clinch, C. Sas, M. Langheinrich, G. Ward, and A. Schmidt. Security and privacy implications of pervasive memory augmentation. *IEEE Pervasive Computing*, 14(1):44–53, 2015.

[7] J. De Winter, D. Dodou, and P. Hancock. On the paradoxical decrease of self-reported cognitive failures with age. *Ergonomics*, 58(9):1471–1486, 2015.

[8] S. Fuchs, K. S. Hale, and P. Axelsson. Augmented cognition can increase human performance in the control room. In *Proceedings of the 8th IEEE Conference on Human Factors and Power Plants and HPRCT 13th Annual Meeting*, pages 128–132. IEEE, 2007.

[9] D. J. Herrmann and U. Neisser. An inventory of everyday memory experiences. *Practical aspects of memory*, 2:35–51, 1978.

[10] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. Sensecam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*, pages 177–193. Springer, 2006.

[11] H. V. Le, S. Clinch, C. Sas, T. Dingler, N. Henze, and N. Davies. Impact of video summary viewing on episodic memory recall: Design guidelines for video summarizations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4793–4805. ACM, 2016.

[12] D. A. Norman. Design rules based on analyses of human error. *Commun. ACM*, 26(4):254–258, 1983.

[13] V. Pejovic, N. Lathia, C. Mascolo, and M. Musolesi. Mobile-based experience sampling for behaviour research. In *Emotions and Personality in Personalized Services*, pages 141–161. Springer, 2016.

[14] L. Ramos, E. van den Hoven, and L. Miller. Designing for the other 'hereafter': When older adults remember about forgetting. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 721–732. ACM, 2016.

[15] B. J. Rhodes. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4):218–224, 1997.

[16] E. Richard, V. Billaudeau, P. Richard, and G. Gaudin. Augmented reality for rehabilitation of cognitive disabled children: A preliminary study. In *Virtual Rehabilitation*, 2007, pages 102–108. IEEE, 2007.

[17] A. Schmidt. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing*, 16(1):6–10, 2017.

[18] C. Seim, T. Estes, and T. Starner. Towards passive haptic learning of piano songs. In *Proceedings of the 2015 World Haptics Conference*, pages 445–450. IEEE, 2015.

[19] S. M. Smith and E. Vela. Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, 8(2):203–220, 2001.

[20] W. S. Terry. Everyday forgetting: Data from a diary study. *Psychological reports*, 1988.

Authors

Sarah Clinch

Dr. Sarah Clinch is a computer science researcher and lecturer at the University of Manchester, UK. She has a PhD in Computer Science from Lancaster University. Sarah's research interests include the application of technology to human cognition and mental health, pervasive display deployments, and privacy and personalisation in ubiquitous computing systems. Sarah sits on the editorial board of IEEE Pervasive Computing and the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), and is an inaugural member of the ACM Future of Computing Academy. Contact her at sarah.clinch@manchester.ac.uk.

Cecilia Mascolo

Cecilia Mascolo is a mother of a teenage daughter. She is also Full Professor of Mobile Systems in the Computer Laboratory, University of Cambridge, UK, a Fellow of Jesus College Cambridge and a Faculty Fellow at the Alan Turing Institute for Data Science in London. Prior to joining Cambridge in 2008, she has been a faculty member in the Department of Computer Science at University College London. She holds a PhD from the University of Bologna. Her research interests are in human mobility modelling, mobile and sensor systems and networking and spatiotemporal data analysis. She has published in a number of top tier conferences and journals in the area and her investigator experience spans projects funded by Research Councils and industry. She has received numerous best paper awards and in 2016 was listed in "10 Women in Networking/Communications You Should Know". She has served as organizing and programme committee member of mobile, sensor systems, networking, data science conferences and workshops. She has delivered a number of keynote talks at conferences and workshops in the area of mobility, data science, pervasive computing and systems. She sits on the editorial boards of IEEE Pervasive Computing, IEEE Transactions on Mobile Computing, ACM Transactions on Sensor Networks and ACM Transactions on Interactive, Mobile, Wearable and Ubiquitous Technologies. More details at www.cl.cam.ac.uk/users/cm542.