Technology-Induced Human Memory Degradation

Document Version
Final published version

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
INTRODUCTION
Humans have long used technologies to extend aspects of their memory [Figure 1]. For as long as such technologies have existed, there have been critics insisting that these augmentations will make humans lazy, leading them to use the technology instead of their brains (e.g. [1]). Indeed, some recent evidence suggests that technologies such as digital cameras [4, 8], social media [8], search engines [7, 14] and digital file storage [7] are negatively impacting our ability to remember.

Recent developments in passive sensing, storage, machine learning, and in wearable actuators and displays have led researchers to express new visions for human memory augmentation [2]. The reach of these envisioned systems go far beyond existing tools to act as integrated “memory prosthetics” that are as ever-present as our companion smartphones, but that provide an even more seamless experience – transparently augmenting our cognition [6]. Such technologies have obvious benefits, improving recall whilst simultaneously freeing up cognitive resource to be allocated to other tasks. However, these technologies introduce a new point of failure – individuals dependent on a technology may become dysfunctional when it is removed. In the best case, individuals revert to their natural state, losing any “superhuman” abilities enabled by the technology. In the worst case, the technology...
has irreversibly altered cognition, leading to poorer performance if the technology becomes absent than if it had never been present. In the case of memory, this may manifest as an absence of memory, or as a distortion in remembered events, skills or knowledge.

In this paper, we explore possible futures for memory technologies, highlighting a need for further research into their use. We propose a “white hat” approach to this research, that seeks to deliberately degrade and manipulate cognition, in order to understand and mitigate potential risks.

A VISION OF THE FUTURE

Jeffrey has long been a keen adopter of technology. His first mobile phone (2001, a Nokia 3310) took the pain out of remembering the phone numbers of family and friends, and his Google calendar quickly became a relationship-saving record of birthdays, anniversaries and other important events.

It’s the summer of 2020 and Jeffrey is sorely tempted to purchase the new MemIt device. Launched by startup MEMIN in February the previous year, the device has catapulted its creators into the public eye; a popular celebrity is rumoured to have invited them to holiday with them! MemIt has been featured heavily at tech expos in the last twelve months, and all the hype suggests that the device transforms you into the best version of yourself. “You are your social network!” declare adverts for the new budget model, promoting the device’s ability to help restore memories of even the most distant of acquaintances.

Two months on, and Jeffrey continues to be amazed by his vastly improved memory – he is sure that his recall now surpasses that of his early 20s. Just last week Jeffrey reminded his boss of a passing comment made during a meeting shortly after he bought his new MemIt – it turns out that the flippant suggestion made that day could actually be the perfect approach for their current client. During his lunch break, Jeffrey heads out to a nearby park. On the way MemIt reminds him of a previous trip made on one of the hottest days of the year; Jeffrey makes a quick detour to pickup a Snup soda to accompany his lunch.

Just before Christmas, Jeffrey’s company announces a set of redundancies. Jeffrey’s job is safe, but his close friend Alan is losing his job. Jeffrey takes Alan out for a drink, but after a few beers Alan observes that none of the redundant staff have a MemIt device despite their growing popularity at the firm. Jeffrey is hurt by the suggestion that MemIt is the reason he has kept his job, he leaves the bar soon after.

A year on, and Jeffrey has never been more pleased with his MemIt. He can think of at least three occasions today when the device has helped him avoid making a foolish mistake. This morning his boss asked about an email sent last week; Jeffrey is sure he’d not seen an email, but MemIt was quick to find the message so that he could provide a quick verbal response with the promise of something in writing later. Thank goodness MemIt is so reliable. Last week’s brief outage left Jeffrey utterly lost (and almost certainly lost his company a new client); he was so glad that MEMIN restored service within just 25 minutes.

Discussion

Just one possible scenario, Jeffrey’s encounters with the MemIt device highlight a number of potential negative consequences for memory technologies.
Two months after his initial purchase, Jeffrey’s experiences seem entirely positive. However, the reader should note Jeffrey’s assimilation of the device and its capabilities into the beliefs he holds about his own faculties. As individuals today have a tendency to misattribute information gleaned from the Internet to their own knowledge and intelligence [3, 11], so too is Jeffrey “amazed by his vastly improved memory.”

On that summer day, we see an indication that the MemIt device has the potential to shape Jeffrey’s behaviour. In this case, something the memory triggered by his device leads Jeffrey to make an unplanned purchase. The (mis)use of memory technologies for advertising has previously been proposed in [2]. In their more innocuous example, an individual is simply reminded of memories associated with a buying need when approaching a store that could fulfil that need; in a more sinister case, the authors note how psychological phenomenon such as retrieval induced forgetting might be used to amplify some memories and attenuate others – potentially allowing big brands to pay to shape the MemIt retrieval algorithm and thus Jeffrey’s memory for products and product experiences.

Later, we see how prosthetics become so ingrained into both personal and social perceptions of ability, and how critical they can become to workplace efficiency that those without devices are left on the wrong side of a new digital divide. It’s not known whether Alan chooses not to use a MemIt device or whether his circumstances prevent use (e.g. because he cannot afford one), but it is highly likely that the emergence of future memory prosthetics could amplify existing social tensions as well as education and class divides.

Towards the end of the scenario we see a case in which the MemIt device is at odds with Jeffrey’s own recollections, leading him to capitulate to the device. Whilst this could simply be a demonstration of the superiority of the device over Jeffrey’s memory, it’s equally possible that technologies could convince their users of the validity of false memories – could it be that the email wasn’t sent when claimed, but data simply added retrospectively? If the evidence is there, the human mind may be convinced (as demonstrated through false memory studies using digitally manipulated photographs [10]).

Finally, we see a hint of the problems caused when a technology on which Jeffrey is now highly dependent becomes unavailable, even temporarily. Whilst transient failures are likely to be the cause of only minor annoyance, this moderate outage of 25 minutes meant that Jeffrey found himself unable to complete work tasks to satisfaction, leading him to have “almost certainly lost his company a new client”. Were the technology to disappear more permanently (e.g. should the MEMIN company go bust), Jeffrey and his contemporaries might be at risk of losing a lifetime of personal memories in addition to the knowledge on which their jobs depend.

A CALL FOR “WHITE HAT” HCI RESEARCH

Recent research has begun to explore negative cognitive impacts of the Internet and smartphones, two of our most dominant technologies [5, 7, 8, 12, 13]. As technology begins to deliberately target
Technology-Induced Human Memory Degradation

There is a growing need for a critical evaluation that focuses on unintended side effects. Research is needed to explore interactions between cognition and technology for two distinct cases:

1. Augmentation systems that are primarily benevolent in nature, but that pose a risk to cognitive function, and
2. Augmentation systems that deliberately manipulate cognition.

In the first case, the core focus of research may simply be the provision for measurement of cognitive function, helping users to understand the unintentional tradeoffs they may be making. Initially, this may take the form of a general disclaimer or warning, but over time systems could provide self-test capabilities and detailed user feedback – “your use of XXX software coincides with a 5% increase in prospective memory, but a 0.34% decrease in phonological working memory.” Key research challenges likely centre on cognitive measurement: which capabilities should be monitored, what form and frequency of testing is appropriate, and how results should be interpreted and delivered to users.

The second class of memory augmentation pose a greater challenge. Inspired by the security domain, we propose a need for “White Hat” or “ethical” memory augmentation research that explicitly sets out to discover and exploit vulnerabilities in human cognition.

The idea of cognitive vulnerabilities as a security concern for human memory was first proposed by Davies et al. [2], with the suggestion that real-time monitoring, similar to current anti-virus software, might be a necessary addition to augmentation technologies. Whilst we agree that safeguards should be added to our emerging cognitive technologies, we go one step further with the security metaphor and identify a need for efforts that build our knowledge of cognitive vulnerabilities in human memory (c.f. the virus definitions of common anti-virus software).

Conducting research in this space raises considerable ethical concerns; deception of human participants is a given (e.g. [10]), and the very need for this research centres on our uncertainty of both short- and long-term effects. One option may lie in the development of cognitive simulations as a safe testbed for attacks. However, given the important role of humans in these vulnerabilities, it’s likely that research will still require human subjects. A middle-ground may see simulations used as a low-fidelity platform against which a battery of attacks can be trialled; once a potential weakness is found, a minimally-damaging corollary of the simulated attack could be proposed for human testing. Following identification of a vulnerability, the next challenge centres on the creation of ‘definitions’ to be used by detection and intervention systems. This requires (1) a need to establish reliable and measurable physiological or behavioural indicators of the ‘attack’; (2) a need for active detection of target indicators; (3) identification and deployment of effective countermeasures. Finally, we note that (as in traditional cybersecurity research) uncovering risks without solutions could lead to a window of opportunity for malicious parties. We therefore envisage the emergence of an ethical (or legal) framework that governs white hat research in the cognitive domain, enabling effective countermeasure deployment before publication of an attack.
REFERENCES


