Prisoner of Words: Lessons Learnt from Mobile Gamification of Lab Memory Experiments

Passant ElAgroudy
LMU Munich
Munich, Germany
passant.elagroudy@gmail.com

Pedram Khoshdani
JDisc GmBH
Gärtringen, Germany
pedram.khoshdani@gmail.com

Tilman Dingler
University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

Geoffrey Ward
University of Essex
Essex, United Kingdom
gdward@essex.ac.uk

Paweł Woźniak
Chalmers University of Technology
Gothenburg, Sweden
pawel.wozniak@chalmers.se

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

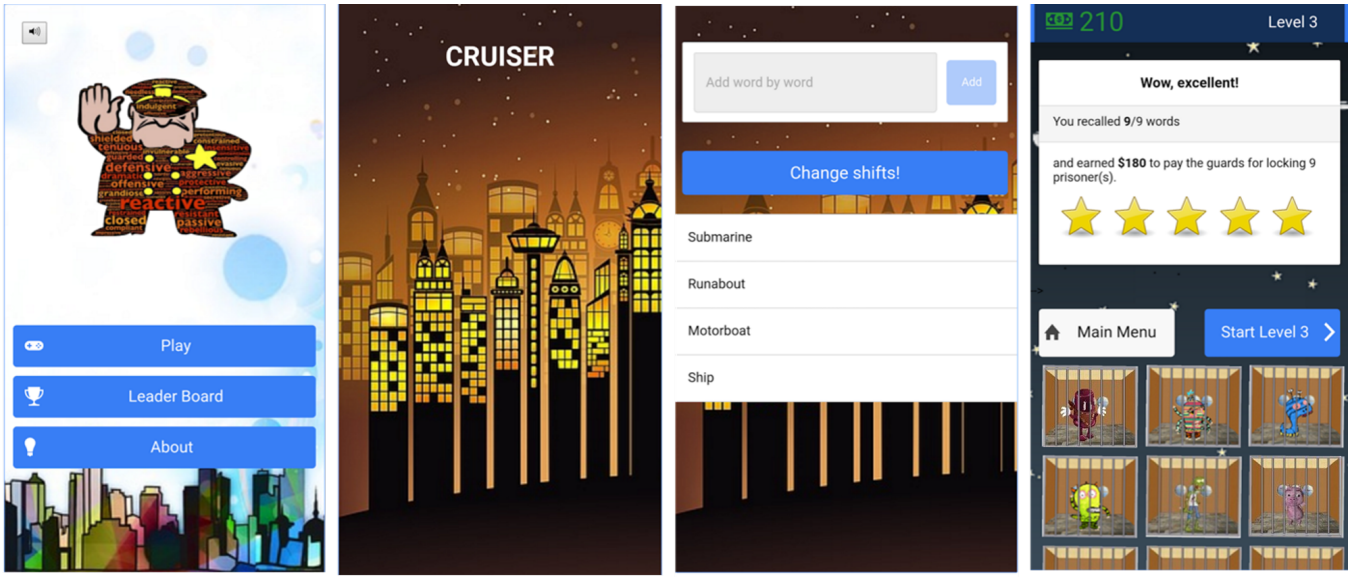


Figure 1: Screenshots showing the game play from "Prisoner of words" memory game replicating a part of Ward et al.'s laboratory memory experiment [17]. The player wins by recalling as many words as possible correctly from a list of words. The player sees one word at a time for one second. Next, (s)he enters as many words as they can in any order and then the results of the recall are shown. Each correct word corresponds to credit to get a prisoner out of their cell.

ABSTRACT

Gamifying experiments in-the-wild showed promising opportunities in supporting participants' engagement, increasing sample size and diversity, and saving lab and personnel resources needed for lab experiments. However, transforming memory lab experiments to mobile games is challenging as some standard game design guidelines jeopardize the validity of the experimental design. Our work draws attention to this trade-off by providing design guidelines for critical game elements, namely scoring systems and input methods. We distil those guidelines from a case study where we replicated a lab memory experiment via a mobile game apparatus and produced congruent psychological results. Our work sheds the light on the special challenges in gamifying memory experiments and encourages psychologists and game designers to investigate human memory in natural settings.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI PLAY '21, October 18–21, 2021, Virtual Event, Austria

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8356-1/21/10.
https://doi.org/10.1145/3450337.3483502

CCS CONCEPTS

• **Human-centered computing** → **Field studies**; **Mobile phones**.

KEYWORDS

memories; in-the-wild deployments; mobile games; design space; lab experiments; serial recall

ACM Reference Format:
Passant ElAgroudy, Pedram Khoshdani, Tilman Dingler, Geoffrey Ward, Paweł Woźniak, and Albrecht Schmidt. 2021. Prisoner of Words: Lessons Learnt from Mobile Gamification of Lab Memory Experiments. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '21), October 18–21, 2021, Virtual Event, Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3450337.3483502

1 INTRODUCTION AND RELATED WORK

Memories influence our emotions and how we perceive individuals and concepts. They contribute to learning and decision making shaping our characters. Therefore, there has been extensive psychology and Human-Computer Interaction (HCI) laboratory literature uncovering how our memory works (e.g. see [2, 3, 5, 6, 15]). With the prevalence of ubiquitous technologies, mobile games offer promising opportunities for training and conducting in-situ memory experiments outside the lab (e.g. [4, 8, 9, 11, 16]). We find the in-the-wild gamification approach promising specifically in the context of memory research because the collected data is a closer approximation for natural absent-mindedness[1] from interruptions while being augmented with contextual cues such as location and weather to support holistic interpretations. Mobile apparatuses like games can save lab and personnel resources [7]. Additionally, they facilitate participants' recruitment despite the high drop out rates [7] as even shortly trying an application can be sufficient to collect data for all conditions in some experiments. Moreover, they enable researchers sometimes to collect numerous encounters of rare conditions. For example, Stephan Mitroff, the designer of *Airport Scanner* game [10] explains "When a target appears only 0.1% of the time, you need 1,000 trials to get a single case. Too many trials are realistically needed to assess the detection of such targets in a lab setting. Using the game, we were able to look at hundreds of cases for each of the nearly 30 rarest targets"[1]. They also support research replicability as tools for on-going data collection.

Despite the advantages of in-the-wild game deployments, transforming a laboratory memory experiment into a game is a challenging task. The interrupted incomplete nature of the in-situ data [7] adds complexity to the data analysis. This is because memory experiments traditionally employ within-subject designs with numerous conditions and trials to avoid individual differences in the recall capabilities. Additionally, human memory is prone to suggestability, cuing, and misattribution. Thus, designers have to carefully iterate and rethink game design standards for fear of accidentally introducing confounding variables to the psychological experiment. Our work here aims to *raise the awareness* of the HCI community about this specific problem.

We achieve this goal by reporting on the collaborative design process in an interdisciplinary team of psychologists and computer scientists to transform an existing laboratory experiment [17] in to an in-the-wild mobile game apparatus. The experiment investigates the effect of a list's length on the recall. The experiment served only as a use case to facilitate the discussion around a common project between the computer scientists and the psychologists so that we can identify such problematic game design elements. The definition of "problematic" in this context is design decisions that can tamper with natural recall. We chose to replicate the lab experiment to gauge the validity of our design recommendations by comparing our collected data trends to existing data in literature. Our approach succeeded as we were able to reproduce the data trends from the original experiment using our mobile apparatus within a short time frame (2 weeks, n=44). Our contribution in this paper is *two-fold*: (1) highlight the the potential adverse impact of competitive scoring and smart input methods on the validity of the experimental data, and (2) replicate and validate Ward et al.'s experiment [17] in the wild with conforming results to the original experiment showing that participants recall shorter lists in forward order while they recall longer lists in backward order. Our work shall aid game designers and psychologists investigate human memory in natural settings without jeopardizing the quality of the collected data.

[1]People suffer from natural gaps in capturing information due to lack of attention [14]

2 METHODOLOGY

In this section, we introduce the original lab experiment [17] and highlight the snippet we replicated and the alterations we did to fit our context. Afterwards, we introduce the game concept and report on the collected data. Next, we report on the participants' distribution and average game play. Lastly, we report on how we analyzed the data.

2.1 Experimental Design

We replicated an existing laboratory memory experiment as a sample use case. We opted for replication rather than designing a new experiment to have a ground truth for the expected results and to focus on the transformation aspect rather than the experimental design aspect. Thus, we *modified* a laboratory memory experiment by Ward et al. [17] which investigated the effect of a list length on the recall quality. Ward et al. had found that as the list length increases, the natural order of recalling the words changes. Participants tend to recall short lists in forward order starting with the first word but they tend to recall longer lists with one of the last few words. This finding could be used to decide the preferred order in lists of information snippets. An example application in Human-Computer Interaction (HCI) is optimizing advertisements' order on an ambient display to maximize memorability. Another example is ordering posts in a news feed as similar recall patterns are observed there [12].

In the lab experiment, the independent variable was the number of words in a list (i.e. *list size*) and the dependent variables were *recall accuracy* and *recall order*. The experiment had fifteen conditions (list sizes= [1,15]). In our game, we quantized the conditions equidistantly to four discrete values (list sizes = 3,6,9,12) and reflected on the recall order only. The objective was to reduce data collection time as our main objective is reflecting on the design process rather than replicating the experiment. We used the same word bank from the original lab experiment.

2.2 Apparatus: Game Design

Our game is a one-player game. Figure 1 shows a summary of the main screens in the game. The game storyline is inspired from a folkloric iranian story book called *Shahname* (meaning "the book of kings"). In the book and the game, a hero called Rostam attempts to save the nation by rescuing the king from the demons. The player

is the warden / supervisor of the prison where the king is captivated. The warden authenticates the identity of the changing prison guards through passcodes formed of a list of words. The warden, i.e. our player and participant, gets a list of words representing the passcodes and has to remember them correctly afterwards, otherwise the guard misses their shift and the king along with the other prisoners can flee. Each passcode, i.e. word, corresponds to a cell with a prisoner.

Level Design. Each level covered all conditions as it had four lists shown at random order using round robin randomization[2]. The starting condition of different players was pseudo-randomized to cover potential combinations. Similarly, the conditions' order per player in each level was pseudo randomized to reduce order bias. Additionally, we were not able to bound each level to a particular list size to avoid learning bias and order effects in the data collection.

Experimental Task. The participant gets a sequence of words in a list. Each word is displayed for one second similar to the laboratory experiment then disappears allowing the next word to show. After the list is finished, the player is asked to type all the words they saw in any order. The player submits the words from each list using the "change shifts" button. We do not mention the list size prior to flashing the words as it could cue the participant about the words they missed.

Input Method. The player types the words. However, we disable the auto-complete and the auto-correct features as they can cue the participant. For example, consider the participant remembering the word "cat" and typing it but the auto complete shows the correct word from the list that is "catch". To account for false negatives from spelling mistakes, we used a correction algorithm, namely Hamming distance algorithm with index number= 1.

Scoring System. After the player submits a list, they get feedback about their performance. The feedback consists of: (1) amount of earned money, where every correctly recalled word corresponds to money given to the guard, (2) number of correctly recalled words compared to the list size (e.g. 8/9 words recalled correctly), and (3) number of cells with prisoners where every forgotten word corresponds to an empty cell. Figure 1 shows the view. The money factor[3] does not change within the same level, i.e. we do not incentivize recall of words in longer lists over shorter lists. This is to avoid tampering with the player's attention and/or recall between conditions. However, the money factor increases between levels to encourage players to continue playing and incentivize data collection. For example, a correct word would earn the player 10 points in level 1 and 40 points in level 4. ($MoneyEarned = CorrectWords * (CurrentLevel * 10)$)

Implementation and Data Collection. The game was developed as a cross-platform mobile game using Ionic Framework. The data was stored locally and synchronized whenever Wifi connection was found to enable playing in internet-less opportune moments. The data was transmitted to a NodeJs Express server stored on the institute's secured servers. The data was modelled using a MySQL database. We collected the following data about players: gender, occupation, country, mobile type, words displayed to them, current level, words they recalled, order of words they recalled.

[2]This is the random selection and elimination of a condition from a pool such that a condition is not repeated twice till all conditions are covered

[3]Earned money per word

2.3 Procedures

Participants were presented a consent form the first time they opened the game. We used an opt-in approach rather than an opt-out approach recommended by [7] to conform with the ethical requirements of informed consent as memory research is considered sensitive. It explained the data we are collecting and informed them that the data is being used for research. We also collected basic demographics as explained in the previous section such as gender, occupation, country and automatically logged the mobile type. Afterwards, participants were allowed to play as much as they want. They were not instructed to finish a minimum number of levels nor to remove the application after a certain period. We automatically synchronized the collected data upon the availability of WiFi connection.

2.4 Participants and Recruitment

We recruited participants through the word of mouth, distributing the game in the researchers' social circle, snowball sampling and publicizing the game on social media. 72 participants played the game. We excluded 28 players who either did not complete the first level, i.e. they missed conditions or failed to recall at least one word in the first level and had no further levels. This indicates that they were not playing the game but only opened to check it. Thus, we analyzed the data from 44 participants (22 males and 22 females, 38% drop rate of participation). They came from eight countries (Iran, Bangladesh, Malaysia, Georgia, France, Germany, Sweden, and United States of America). The mean age was 29 years old (minimum = 15 and maximum= 54). Participants submitted an average of 14.45 lists (SD= 16.6). The data collection lasted for two weeks. We did not incentivize the participation for the study.

2.5 Analysis Methods

To validate the collected data trends, we used similar techniques from the lab experiment to analyze the game data. We divided each list into three equal intervals. Afterwards, we labelled the trials to indicate the interval of the first recalled word based on it's serial position in the list. The labels were called *recall category*. The labels were used to understand whether participants tend to recall the lists in forward or backward order. Words from the first interval were labelled as *FIRST*, words from the second interval were labelled as *OTHER*, and words from the last interval were labelled as *LAST*. Table 1 summarizes the position of the first recalled word in a list and the corresponding label it gets based on the list length. For example, a trial where the the sixth word on a 6-word list is recalled first will be labelled as LAST. However, a sixth word on a 9-word list is labelled as OTHER.

Afterwards, we counted all the trials of a particular list size for a player. We used the total number of trials to calculate the probability of a participant recalling one of the categories. We also calculated the number of trials where participants did not recall any word.

Table 1: Summary of the serial position of the first recalled word in a list and its corresponding recall category. A list is divided into three equal partitions. FIRST indicates a forward recall list. LAST indicates a backward recall. OTHER indicates a list where a participants starts by recalling its middle element.

List size	FIRST	OTHER	LAST
3 words	1	2	3
6 words	[1,2]	[3,4]	[5,6]
9 words	[1,3]	[4,6]	[7,9]
12 words	[1,4]	[5,8]	[9,12]

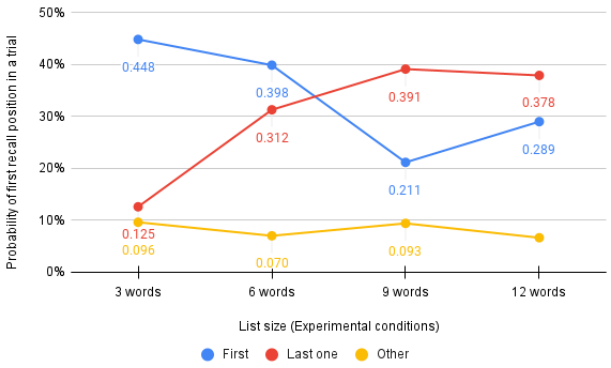


Figure 2: Probability of the serial position of the first recalled element in a list. This indicates whether participants are likely to recall lists from the start or the end. In accordance with the lab experiment by Ward et al. [17], participants recall shorter lists from the start (size=3 and 6). However, they recall longer lists starting from the end (size= 9 and 12).

3 RESULTS: GAME AND LAB EXPERIMENT CONFORM

Figure 2 summarizes the descriptive results. The data trends indicate that participants recalled shorter lists (list size=3,6) in forward order, i.e starting by the first word. However, participants recalled longer lists (list size= 9, 12) starting in backward order, i.e. starting by the last words. The results are consistent with the lab experiment from Ward et al. [17].

The probabilities are calculated with respect to the total number of trials of a certain "List Length (LL)" per participant. This implies they also include zero recall trials where participants did not recall any words. Participants recalled no words in 33.12% of the trials in with LL= 3, followed by 30.5% of the trials in LL= 9, 26.63% of the trials in LL= 12, and 22% of the trials in LL= 6. Table 2 summarizes the probabilities of recall after excluding the zero-recall trials. The trends are the same as the ones presented in Figure 2.

Table 2: Probability of first item recall position in a list based on its size. The probabilities are calculated after removing zero-recall trials. The trends conform with the ones presented in Figure 2

Recall starts with	LL=3	LL=6	LL=9	LL=12
FIRST	**57.11%**	**45.08%**	27.76%	34.37%
OTHER	13.57%	12.18%	11.17%	9.21%
LAST	18.51%	39.44%	**55.82%**	**52.98%**

4 LESSONS LEARNT AND TAKEAWAYS

In this section, we reflect on our experience while designing the game. The distilled points are extracted from the team's interdisciplinary arguments between psychologists and computer scientists to reach design consensus. We report on aspects that could be useful to other game designers beyond the scope of this particular experiment. However, we make no claims about the generalizability of the results but rather nudge the reader to notice the design problem.

4.1 Merits Can Interfere with Recall

Score is a classic engagement element in game design. However, competitive reward strategies in memory apparatuses could tamper with the experimental validity of the data. Rewarding actions motivate players to remember better (e.g. [13]) and to possibly perform drastic measures to gain points. An example is the player recording then copying stimuli into the game rather than relying on their personal recall. Lack of reward systems though does not provide sufficient gratification to play the game, i.e. participate in the experiment. Thus, we have *three* design recommendations:

(1) **Avoid rewarding secondary actions** We recommend focusing on a simple reward system for the main experimental task only rather than adding complexity to the game play through secondary rewardable actions. For example, a common method in game design to increase game complexity is to use performance time as a scoring factor. However, in some memory experiments like ours, rewarding faster responses can unnecessarily increase the error rate and impact the recall rate due to stress.

(2) **Wisely use real-world merits** Physical merits such as vouchers and money for top-performing players can encourage fast collection of larger datasets. However, it could also negatively impact the quality of collected data if players tamper with the results or introduce drastic actions such as recording to "win". Translating real-world norms for rewarding participation time could also be problematic for data integrity. For example, rewarding the player who played the longest time could correspond to mindless submission of stimuli, i.e. low quality data.

(3) **Report on the memory task** We suggest reporting the results of the memory task rather than an abstract score to enhance the player's sense of progress, achievement and contribution to the research. For example, if the task is to remember five words and the player remembered three, the system message could be (*you remembered 3 / 5 words and earned 30 points*) rather than the earned points only.

4.2 Smart Input Methods Can Interfere with Recall

Controls with predefined options like autocomplete text fields and dropdown menus are better suited for recognition rather than recall tasks as they cue the player to remember the stimuli. Thus, they tamper with the experimental data. However, using basic versions (e.g. textfield without autocomplete) jeopardizes the user experience because it is cumbersome. It also increases the error rate due to human error during data entry rather than recall failures. We recommend the following *two* design guidelines:

(1) **Choose convenient stimuli** The complexity of the chosen stimuli such as number of stimuli and length of words or number of details in a picture should be controlled based on the input method to minimize player's frustration and data entry error rates. For example, in our scenario, we specifically chose shorter words from the dataset as we used a standard text field with no prediction features.
(2) **Design error recovery strategy** There are two approaches to overcome the loss of data points due to entry errors rather than memory failures. The first is to increase the error tolerance to account for cases such as typing "chiar" instead of "chair". The second is to enable editing of entries. Designers should cautiously use the latter approach. Despite reducing false negative hits, it alters stimuli recall order as players go back and forth between entries which might not be favourable in an experiment about serial recall order for example.

4.3 Overcome Missing Conditions through Level Design

Memory experiments usually employ a within-subject design to neutralize personal recall capabilities. Conditions in a memory task usually have a natural incremental complexity such as number of stimuli. An example is if the independent variable is list size, it is harder to remember a list of 12 words compared to a list of 3 words. Within-subject designs are challenging in mobile deployments due to the unpredictable high dropout rates [7] which can lead to excluding large number of participants because of incomplete datasets. Levels in game design is a good tool to encourage users' engagement by giving a sense of progress and increasing game complexity. Thus, we have *three* design recommendations for levels to minimize incomplete datasets:

(1) **Bundle conditions in a level** In simple games with shorter levels, users usually play several levels regardless the quality of the game. Possible approaches to designing a level are (1) each level corresponds to an experimental condition and the level contains several trials (game tasks) or (2) each level corresponds to trials from several conditions. We recommend the latter where each level contains at least two trials from all conditions. The reason is to ensure that a player who plays as little as a single level in the game provides viable analysis data. We recommend having two trials per condition in case the researchers drop the first trial as training material.
(2) **Avoid leveling through condition complexity** Researchers should not map conditions' complexity to game levels as it jeopardizes the experimental validity. This is relevant where conditions are inherently ordinal such as increasing the number of words or list length. For example, we avoided matching the incremental conditions to the levels in our game where level 1 is 3-word lists then level 2 is 6-word lists. This grouping would have created a learning bias and introduced order effects to collected data.
(3) **Use step conditioning techniques** We propose mapping the levels to subsets of the conditions and collecting them incrementally spanning the conditions' spectrum in case of ordinal conditions. For example, assume we want to conduct our experiment but on 8 list sizes =$[3, 10]words$. A possible design is to have levels 1 and 2 correspond to trials with list sizes = 3, 5, 7, 9 words then levels 3 and 4 correspond to trials with list sizes = 4, 6, 8, 10 words. This ensures that at least all players submit data to coarsely-sampled conditions instead of losing all the participant's data points if they drop in the middle of a long level having the eight conditions. The conditions' order within the level should be randomized for each player to avoid order effects.

4.4 Design for Shorter Usage

As we know from prior work that users shortly use mobile apparatuses and it is critical to collect the required information as fast as possible [7], we recommend designing games with simple, short, and repetitive tasks. This can promote using the application in opportune moments such as waiting for bus.

5 CONCLUSION AND LIMITATIONS

In a spirit of reflective practice, we contribute insights for future research at the intersection of experimental psychology and game design in HCI. In this paper, we highlight the potential negative impact of designing competitive scoring systems, complex levels and smart input methods on the validity of memory experiments performed by serious games. We propose design guidelines to overcome such challenges through reflecting on a case study of transforming an existing laboratory study to a deployed mobile game. We found that scoring can over-incentivize players to artificially augment their recall. Additionally, smart input methods like autocomplete cue recall interfering with the natural errors and serial order. We acknowledge that currently we do not empirically prove our design recommendations or distill the impact of each element alone on the recall. However, the conforming results between our quick game deployment and the original lab experiment after applying the proposed guidelines is promising. Thus, our aim is to: (1) encourage designers to evaluate their game design strategies early on with psychologists in interdisciplinary teams and for both parties to mind the semantic differences in design priorities and constraints, and (2) encourage future systematic investigations quantifying the impact of such widely-used elements on experimental data. We also acknowledge that the high dropout rates might indicate that the game is not enjoyable for longer times. However, it is worth noting that the rates are lower than literature standards (e.g. ca. 40-50% in [7]). Additionally, this showcases that this might not problematic in basic memory research as even shorter sessions provide

enough data for patterns and significantly saves the time of personnel. However, for experiments requiring large number of trials, we recommend investing time in enhancing engagement metrics. Our findings show the feasibility of using serious games as tools for continuous validation and replication of memory experiments in natural settings.

ACKNOWLEDGMENTS

This work has been partially conducted within the Future and Emerging technologies (FET) programme with the 7th Framework programme for Research of the European Commission, under FET grant number:612933 (RECALL) and within the European Union's Horizon 2020 research and innovation programme, under ERC grant number: 683008 (AMPLIFY).

REFERENCES

[1] [n.d.]. Crowdsourcing with Mobile Apps Brings 'Big Data' To Psychological Research. https://www.pharmacytimes.com/news/crowdsourcing-with-mobile-apps-brings-big-data-to-psychological-research. Last accessed: 2021-01-11.

[2] Michael C Anderson, Robert A Bjork, and Elizabeth L Bjork. 1994. Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 5 (1994), 1063.

[3] Alan Baddeley, Michael W. Eysenck, and Michael C. Anderson. 2009. *Memory*. Psychology Press Taylor and Francis Group.

[4] Harriet R Brown, Peter Zeidman, Peter Smittenaar, Rick A Adams, Fiona McNab, Robb B Rutledge, and Raymond J Dolan. 2014. Crowdsourcing for cognitive science–the utility of smartphones. *PloS one* 9, 7 (2014), e100662.

[5] Jason R Finley, William F Brewer, and Aaron S Benjamin. 2011. The effects of end-of-day picture review and a sensor-based picture capture procedure on autobiographical memory using SenseCam. *Memory* 19, 7 (2011), 796–807.

[6] Morgan Harvey, Marc Langheinrich, and G. Ward. 2016. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive Mob. Comput.* 27 (2016), 14–26.

[7] Niels Henze, Martin Pielot, Benjamin Poppinga, Torben Schinke, and Susanne Boll. 2011. My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 3, 4 (2011), 71–91.

[8] O. V. Hove, A. V. Muylem, Vasileios Andrianopoulos, D. Leduc, V. Feipel, G. Deboeck, and B. Bonnechère. 2020. The use of cognitive mobile games to assess the interaction of cognitive function and breath-hold. *Respiratory Physiology & Neurobiology* 274 (2020).

[9] A. Koivisto, Antero Lindstedt, S. Merilampi, and K. Kiili. 2016. Designing Working Memory Games for Elderly. In *GALA*.

[10] Stephen R Mitroff, Adam T Biggs, Stephen H Adamo, Emma Wu Dowd, Jonathan Winkle, and Kait Clark. 2015. What can 1 billion trials tell us about visual search? *Journal of experimental psychology: human perception and performance* 41, 1 (2015), 1.

[11] Malin Nilsen, Mona Lundin, Cecilia Wallerstedt, and Niklas Pramling. 2018. Evolving and re-mediated activities when preschool children play analogue and digital Memory games. *Early Years* 41 (2018), 232 – 247.

[12] Pawarat Nontasil and Stephen J. Payne. 2019. *Emotional Utility and Recall of the Facebook News Feed.* Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3290605.3300252

[13] Charlotte Saucet and M. Villeval. 2019. Motivated memory in dictator games. *Games Econ. Behav.* 117 (2019), 250–275.

[14] Daniel L Schacter. 2002. *The seven sins of memory: How the mind forgets and remembers.* HMH.

[15] Abigail J Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do life-logging technologies support memory for the past?: an experimental study using sensecam. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 81–90.

[16] Laura Lim Sie-Yi and Noraziah ChePa. 2020. Criteria of Mobile Psychotherapy Games for Memory Disorder: A Systematic Literature Review. *Int. J. Interact. Mob. Technol.* 14 (2020), 57–72.

[17] Geoff Ward, Lydia Tan, and Rachel Grenfell-Essam. 2010. Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, 5 (2010), 1207.