Serious Game Development for Detecting Mild Cognitive Impairment

by
Kyle Leduc-McNiven

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

Copyright © 2022 by Kyle Leduc-McNiven

Abstract

Mild Cognitive Impairment (MCI) is cognitive decline before one is naturally expected to have age-related cognitive decline but not severe enough that it affects a person's daily routine or life. MCI is often a precursor to the various forms of dementia before the symptoms are manifest and diagnosed. If MCI could be reliably assessed, the prognosis of more serious decline may be improved. This thesis will discuss the development of a serious game denoted War Cognitive Assessment Tool or WarCAT, with the objective being to aid in the early detection of MCI. This thesis will also overview contributions to papers published, as well as the work done on the previous iteration of WarCAT denoted War Data on Gambling (WarDOG) whose purpose is to help in the study of gambling and gaming addictions.

Keywords: mild cognitive impairment, dementia, serious games, technology for early MCI assessment

Acknowledgements

I would like to thank Dr. Robert McLeod for all the guidance and support he has given me throughout my many years under his wing, as well as his help with editing and writing this thesis. I would also like to thank Dr. Marcia Friesen for her aid, support, and guidance over the years. Without the support of both Dr. McLeod and Dr. Friesen none of this would have been possible, so my sincerest gratitude goes out to both for everything they have done. I also am grateful to my colleagues Ryan Dion, Andrew Kozar, Mahmood Aljumaili, Ben White, Genico Melegrito, Belema Jack-bara, and Michael Wurtak, for helping with the development of the game, its features, and its research. A special thanks to Dr. Kai Gutenschwager for his aid in the development of the current cognitive fingerprint. Lastly, I would like to thank my parents, Lucie Leduc and Neil McNiven, for allowing me the opportunity to do this endeavor by, feeding, housing, and helping in all aspects of my life.

Table of Contents

List of Figures & Tables

Chapter 2 figures

Chapter 3 figures

Chapter 4 figures

Chapter 4 Tables

List of Abbreviations

MCI-Mild Cognitive Impairment
AD- Alzheimer's Disease
CDR-Clinical Dementia Rating
GDS- Global Deterioration scale
CIND- Cognitive Impairment No Dementia
MMSE- Mini Mental State Exam
fMRI- functional Magnetic Neuroimaging
PET-positron emission tomography
FDG-PET- fluorodeoxyglucose positron emission tomography
APOE4- apolipoprotein E4
Aβ42 - Amyloidβ 42
T-tau - Tau
P-tau181 - Tau phosphorylated at threonine 181
CSF - cerebrospinal fluid
MoCA- Montreal cognitive assessment
NFC - Near Field Communication
WarCAT- War Cognitive Assessment Tool
WarDOG- War Data on Gambling
PvP - Player versus Player
PvB- Player versus Bot
META- Most Effective Tactics Available
GAN - Generative Adversarial Network
RL - Reinforcement Learning

Chapter 1: Introduction

Cognitive decline is something that affects everyone to a certain degree whether it be as a precursor to more serious cognitive impairments or just via the natural aging process. Mild Cognitive Impairments (MCIs) are typically defined as a decline in cognitive ability greater than what is expected during the normal aging process, according to the Canadian Alzheimer Society [1]. The biggest issue with MCIs is that they are often very difficult to discover and once they are discovered, the impairment has often progressed beyond the stage where one could have more easily dealt with it. The conjecture in this thesis is that there is value in finding protocols or methods to detect MCIs earlier on, so that therapies and treatments can be initiated sooner to delay associated downstream dementias.

The issue with this line of thought however is that there are not very many reliable ways of detecting MCIs at the moment, even within a clinical setting. Thus, collecting data from serious games for a person's game play, decision making, memory and judgment on their game play strategies might play a role in helping to assess cognitive health. In this thesis, a serious game is discussed as a potential and/or partial solution to the problem of helping to detect MCIs. The game developed is called War Cognitive Assessment Tool (WarCAT). WarCAT started as a simple card game resembling the card game WAR [2], where you would be pitted against your opponent in a short and quick battle of wits. The idea is that as a person plays the game, they will naturally develop a strategy for success and the longer they play, the more refined the strategy becomes. As player strategies emerge, it is posited that one could detect signs of MCI with game play inference. To simplify the idea; it is assumed that as someone plays a game, they will naturally develop a strategy for success. However, if they have a cognitive impairment of some kind, they may not learn an effective strategy as quickly as a healthy individual, or may learn an effective strategy and then have difficulty recalling it, or not learn an effective strategy at all, all of which lead to more random game-play. If this conjecture were to be viable, a player's game play strategy would have to be part of a very large data set or record of many different players or many demographics. In an attempt to address this concern, the research focused on means to acquire a large enough set of data samples for comparison and analysis to be meaningful.

Smartphones and tablets were chosen as the best platform for the game due to the fact that nearly everyone, including older adults, has a smartphone or tablet of some kind, and it can be a powerful tool for data gathering. Using the concept of capturing gameplay strategy in combination with a potentially vast volume of data would allow us to establish

baselines and compare people of different cognitive abilities. This would potentially allow us to see how likely a person would be categorized or assessed as potentially suffering from an MCI in comparison to people of similar age, gender, and many other demographic characteristics.

This thesis is organized as follows: Chapter 2 provides a literature review of cognitive health issues with the focus on MCI and dementia. Chapter 3 provides a literature review of serious games and their possible uses. Chapter 4 details the development of both games, WarDOG and WarCAT as serious game data collection vehicles. Chapter 5 is a collection of the research papers that resulted from the game WarCAT developed and discussed in Chapter 4. These are expanded upon in realizing the difficulty of collecting a large enough player base and concomitantly the player data to allow for classification and hence make MCI detection viable through serious games. As a large dataset of player data would be essential in the utility of a serious game, data within an AI, or machine learning environment. This is addressed in the research papers presented in the final chapters of this thesis. The final Chapter summarizes the research and presents ideas for future research.

References

[1] Alzheimer Society of Canada. n.d. “Mild cognitive impairment.” Alzheimer Society of Canada. Accessed February 15, 2022. https://alzheimer.ca/en/about-dementia/other-types-dementia/conditions-related-dementia/mild-cognitive-impairment.

[2] “War – Card Game Rules.” n.d. Bicycle Playing Cards. Accessed March 6, 2022. https://bicyclecards.com/how-to-play/war/.

Chapter 2: Cognitive Impairment Literature Review

Mild Cognitive Impairments

Mild Cognitive Impairment (MCI) has a broad definition that has changed over the years. MCI is commonly referred to as cognitive decline greater than expected for age and education level but not to the degree that it actively interferes with a person's life. MCI can appear in all ages but typically manifests itself later in life from age 65 with an approximately 12-18% rate of appearance [1]. MCI is hard to detect and requires long interviews and tests to determine not only the presence and type of MCI, but the likelihood of progression to other forms of dementia or Alzheimer's Disease (AD). What makes this process more complicated is that the definition of MCI is being updated frequently to better define the various forms. The symptoms according to the Alzheimer's society in Canada are: memory loss, impaired thinking skills, disorientation in both time and space, difficulties with language and speaking, poor judgment, and impaired depth perception [2]. While MCI detection is challenging, as with any form of disease or illness, early detection and assessment is highly correlated with a favorable prognosis.

A Brief History of Mild Cognitive Impairments definition

The first use of the term MCI was in the context of two clinical scales, the Clinical Dementia Rating (CDR) and the Global Deterioration Scale (GDS). These scales were used to determine the boundary between normal age-related cognitive decline and dementia. The CDR scales from 0-1 with 0 being healthy 0.5 being mildly impaired and 1 being mildly demented. The GDS is formed of 7 stages where stage 1 has no cognitive impact or decline, and it scales up from there to stage 7 which is very severe cognitive decline [3]. The first proper definition of MCI is in the GDS at stage 3 which is defined as mild cognitive decline with the clinical characteristics being,

> Earliest clear-cut deficits. Objective evidence of memory deficit obtained only with an intensive interview conducted by trained geriatric psychiatrist. Concentration deficit may be evident in clinical testing. Patients may demonstrate decreased facility in the following: (a) remembering names upon introduction to new people; and (b) retaining information after reading a passage from a book.
>
> Decreased performance becomes manifest in demanding employment and social situations. Examples can include the following: (a) co-workers become aware of the person's relatively poor performance; (b) difficulties in finding words and names may become evident to intimates; (c) may lose or

> misplace an object of value; and (d) getting seriously lost when traveling to unfamiliar locations.
>
> Subtlety of the clinical symptoms may be increased by denial that often becomes manifest with these patients. Mild to moderate anxiety also accompanies the symptoms, typically when a patient is forced to cope with challenging employment and social demands that they find they can no longer negotiate [3].

The key points to note are the need for an in-depth interview to notice the deficits and that diagnosis can become more difficult depending on the patient's thoughts on whether they have MCI or not, as such it can make diagnosis very difficult especially under the GDS definition. There are also some experts that say the GDS definition is lacking for determining the difference between MCI that deteriorates into AD and Dementia and the forms of MCI that are less likely to have severe deterioration [4]. Furthermore, dementia diagnoses are only validated when the clinical professional is certain of the diagnosis or specifically "to cases in which there is clear evidence of progressive and significant deterioration of intellectual and social or occupational functioning" [4]. The definition and the scales mean that there are people who meet multiple criteria for dementia but do not get a dementia diagnosis, and this is where MCI comes into play.

Another term used to represent cognitive impairment before the proper usage of the term MCI is called, Cognitive Impairment no Dementia (CIND), and was used in reference to cognitive impairments that may not progress to dementia. CIND has been used in several studies including a Canadian population-based study [4]. In the CIND definition, the criteria for MCI were expanded from just memory deficits to include many different areas of cognitive deficits and clinical subtypes such as; memory impairment due to drug and alcohol use, psychiatric illness, mental impediment, and vascular pathologies [4]. The definition of CIND by the Alzheimer's Society of Canada has been incorporated into the definition of MCI and the term MCI has been expanded into two subtypes, amnestic-MCI and non-amnestic-MCI.

Diagnosing MCI

There are two types of MCI in current literature, amnestic and non-amnestic. These types are distinguished by whether they affect the person's memory or not. Figure 2.1 illustrates the difference between the two types and how clinicians can go about diagnosing the different types [5]. Each of the types is further broken down by whether they affect multiple cognitive domains or only a single domain. To make the diagnostic process more accurate, there are multiple tests that can be performed to help a clinician

test to see if a patient's cognitive ability is reasonable for their age and education level. The most predominantly used tests are the mini mental state exam (MMSE) which tests more for amnestic MCI, and various other neurological surveys oriented towards partitioning a person's basic memory from function. To further diagnose a patient, a clinician needs to determine whether it is single or multiple domain MCI. Single domain amnestic MCI is solely memory cognitive issues whereas multiple domains, as its name implies, pertains to multiple domains of cognitive impairment such as language, executive function, or visuospatial skills. Non-amnestic subtypes are the same as the amnestic version minus the memory impairment.

Other tests for a better picture of the patient's status include Magnetic neuroimaging or functional MRI (fMRI), positron emission tomography (PET), fluorodeoxyglucose PET (FDG-PET), and quantitative electroencephalography. A study by Petersen [6] went further in depth with their analysis, where they did both cognitive and physical tests; including,

> The Hachinski Ischemic Scale, a test of mental status and a neurological study. The test also obtained personal information about the subjects' living situation including a record of independent living, where they were on the Geriatric depression scale, and a total family history likely to check for signs of cognitive decline and depression. The physical/ Laboratory studies included chemistry group, blood cell count, sedimentation rate, vitamin B12 and folic acid levels, thyroid-stimulating hormone level, and syphilis serological testing [6].

Another paper discussed the major findings and topics of MCI:

> An international working group on mild cognitive impairment formulated specific recommendations for criteria, including: (1) the individual is neither normal nor demented; (2) there is evidence of cognitive deterioration, shown by either objectively measured decline over time or subjective report of decline by self or informant in conjunction with objective cognitive deficits; and (3) activities of daily life are preserved and complex instrumental functions are either intact or minimally impaired.[8] These criteria serve to expand the construct of mild cognitive impairment to involve cognitive domains other than memory and make it a prodrome to multiple types of dementia [4].

Other cognitive domains that need to be looked at are simple in nature but difficult to detect and they can be broken into several major categories; Degenerative, vascular, psychiatric, or secondary to concomitant medical disorders (congestive heart failure, diabetes mellitus, systemic cancer). Figure 2.2 [1] showcases how the subtypes and these categories interact in a very clean and easy to understand manner [1].

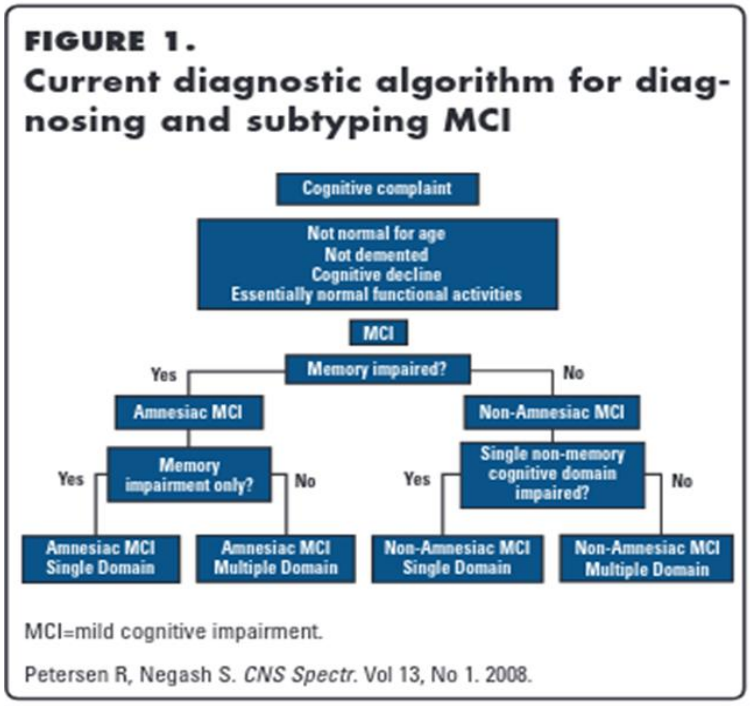


Figure 2.1 Flow chart of diagnosis for the two types of MCI and their subtypes from[5]

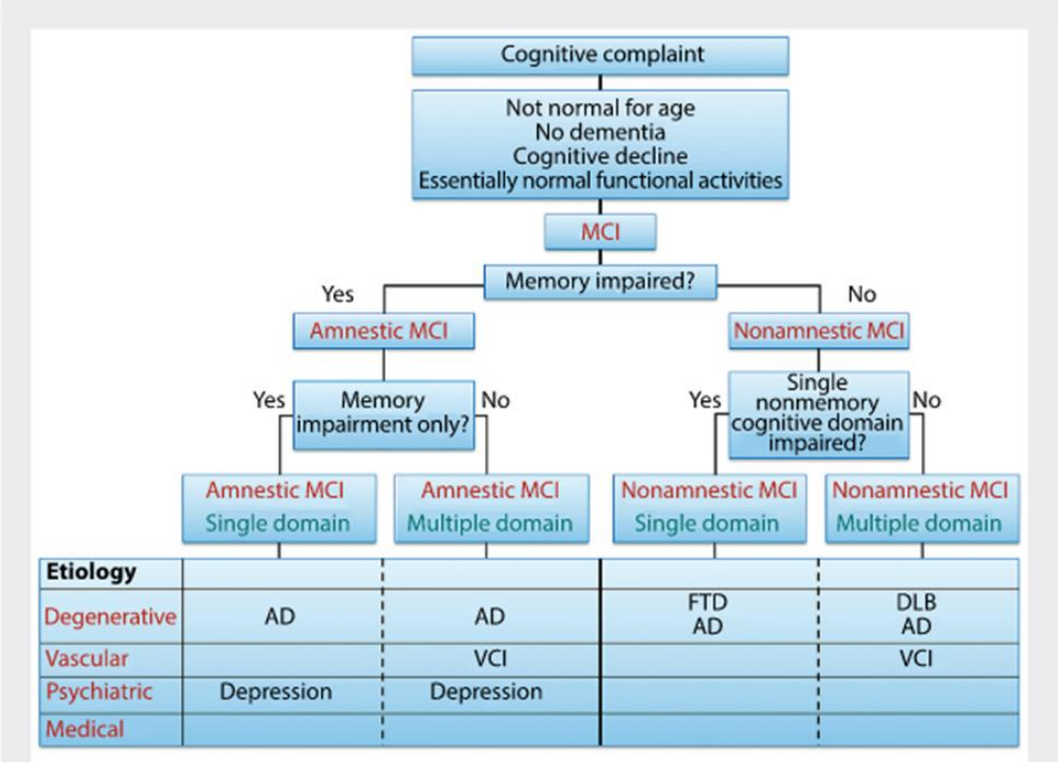


Figure 2.2 A flowchart for diagnosis that includes the subtype chart and their causes [1]

MCI Studies on Progression to Dementia and AD

Most of the research studies on MCI deal with the topic of MCI that progresses to Dementia or AD. The common link between all the studies is that there is no clear factor that determines whether MCI will progress into AD or dementia. One study saw rates of progression to AD at about 5-6% of people after one year. Other studies show similar numbers at around 16% after one year and as high as 64% after having been diagnosed with MCI for more than 2 years without treatment [4]. However, in another study presented in [4], a study was mentioned that is a good representation of how little is truly known about the progression,

> Findings of longitudinal population studies, have shown a prevalence in the general elderly population between 3% and 19%, with an incidence of 8–58 per 1000 per year, and a risk of developing dementia of 11–33% over 2 years. Conversely, findings of population-based studies have shown that up to 44% of patients with mild cognitive impairment at their first visit were estimated to return to normal a year later [4].

Another study [6] shows a very clear gradual decline into AD. The study starts with a large control group of 76 people split into a control of normal people and several people diagnosed with MCI by way of interviews. The study took place over the course of 11 years and showed a steady and obvious decline in cognitive ability at "10-12% per year" [6].

Other studies have taken a more biological approach to finding the link between MCI and AD. Those studies looked at various proteins and physical symptoms to help indicate when a patient has increased odds of progression from MCI to Dementia or AD. Some of the biomarkers for detecting the potential progression are medial temporal lobe atrophy and hypermetabolism, in addition the apolipoprotein E4(APOE4) also acts as a good indicator for progression to AD as those with the genotype tend to progress at a rapid rate. Performing various PET scans including FDG-PET amyloid PET and tau PET can give good insight as well, especially a positive amyloid PET [1]. Amyloidβ 42 (Aβ42), total Tau (T-tau) and Tau phosphorylated at threonine 181 (P-tau181) in cerebrospinal fluid (CSF) have been proven to be effective markers in detecting MCI and its progression to AD [7]. Hansson also claims the rate of progression of MCI patients to AD and Dementia at a rate of 30-50% within 5 years [7]. Using that information, Petersen and colleagues proposed a new flow chart for diagnosis that includes the use of neuroimaging to detect for Aβ42 and tau levels in the patients, as per Figure 2.3 [1].

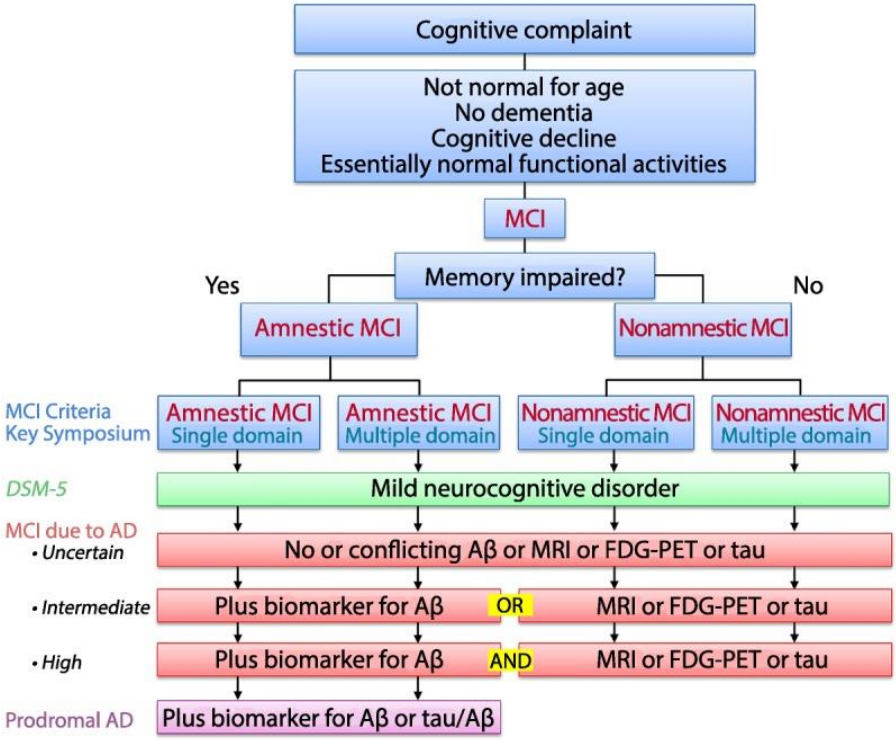


Figure 2.3 Temporal evolution of criteria for mild cognitive impairment (MCI) and prodromal Alzheimer disease (AD). DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; NIA-AA = National Institute on Aging–Alzheimer's Association [1].

Treatments/Management for MCI

The treatments for MCI are very dependent on the cause of the MCI and whether one would be likely to progress to AD or other dementias. The other issue is that depending on the type of MCI, there may not be a treatment method developed yet. For MCI that involves depression or hyperthyroidism, treating those problems themselves will typically solve the problem to an extent provided the issue was dealt with early enough. There are some studies that show certain pharmaceuticals aimed at the APOE4 using donepezil can have positive results, however the sustained benefit is limited [2]. Vitamin E has also been tested against those who had confirmed MCI though the type of MCI was not specified in the paper [5]. Unfortunately, vitamin E was proven to not significantly reduce the risk of progression.

Other methods that could be effective at treating or managing all types of MCI is cognitive rehabilitation [8]. Cognitive rehabilitation utilizes varying strategies, the most interesting of which are computer assisted training programs and more importantly serious games. Some of the serious games that can help with cognitive rehabilitation and

maintenance are games like Lumosity™ [9] and Brain Age™ [10] where the goal is to enhance cognitive ability and help maintain good cognitive health.

Other strategies that have been tested are proper maintenance of physical health, including proper diet and exercise, and reducing stress and anxiety factors [5]. This is thought to be associated with improved health in general and to aid in the rehabilitation of MCI caused by psychological factors like depression, anxiety, and stress. The general consensus in the literature is that there is no clear medication or treatment at the moment that will stop the progression to AD or dementia, but methods to slow the progression can occasionally work, although with mixed results.

One method that can be shown to be both effective in managing cognitive decline and potentially assessing MCI inexpensively as mentioned above is serious games. With the use of serious games, large scale data can be collected and used to establish baselines and then compare individuals. By extension, machine learning can be applied for more detailed analyses and classification. Using mobile serious games would increase the scalability of assessment even further, allowing for a much larger base for comparison. Concomitantly, the scale-up of an engaging and popular serious game would allow for more people to become aware of MCI and its underlying symptoms, risks, and potential treatments.

References

[1] Petersen, Ronald C. 2016. “Mild Cognitive Impairment.” *Continuum*, no. 22, 404-418.

[2] Alzheimer Society of Canada. n.d. “Mild cognitive impairment.” Alzheimer Society of Canada. Accessed February 15, 2022. https://alzheimer.ca/en/about-dementia/other-types-dementia/conditions-related-dementia/mild-cognitive-impairment.

[3] BCGuidleines.ca. 2014. “Global Deterioration Scale.” Global Deterioration Scale. Accessed February 20, 2022. https://www2.gov.bc.ca/assets/gov/health/practitioner-pro/bc-guidelines/cogimp-global-deterioration-scale.pdf.

[4] Gautier, Serge, Barry Reisberg, Michael Zaudig, Ronald C. Petersen, Karen Ritchie, Sylvie Belleville, Henry Brodaty, et al. 2006. “Mild Cognitive Impairment.” *Lancet* 367 (April): 1262-1270.

[5] Petersen, Ronald C., and Selam Negash. 2008. "Mild Cognitive Impairment: an overview." *CNS Spectrums*, no. 13 (January), 45-53.

[6] Petersen, Ronald C., Glenn E. Smith, Stephen C. Waring, Robert J. Ivnik, Eric G. Tangalos, and Emre Kokmen. 1999. "MIld Cognitive Impairment Clinical Characterization and Outcome." *Arch Neurol*. 56:303-308.

[7] Hansson, Oskar, Sylvain Lehmann, Henrik Zetterberg, and Piotr Lewczuk. 2019. "Advantages and disadvantages of the use of the CSF Amyloid β (Aβ) 42/40 ratio in the diagnosis of Alzheimer's Disease." *Alzheimer's Research & Therapy* 11 (April). https://alzres.biomedcentral.com/articles/10.1186/s13195-019-0485-0.

[8] Petersen, Ronald C. 2011. "Mild Cognitive Impairment." *The New England Journal of Medicine* 364, no. 23 (June): 2227-2234

[9] "Lumosity." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Lumosity.

[10] "Brain Age." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Brain_Age

Chapter 3: Serious Games Literature Review

Serious Games are a sub-genre of games that are designed not only for entertainment but for some other primary purpose. Serious games have been around for almost as long as games have existed, with one of the earliest known examples being a game from 1400 BC called Mancala, which was used to teach accounting and trading [1]. Serious games have developed into a flourishing industry, with the industry reporting about 10 billion Euros in revenue in 2010, and the numbers of games in research and industry are rising steadily [1]. Figure 3.1 shows the growth rate of serious games in industry and research up until a decade ago [1].

However, Google trends for the topic of serious games shows interest in industry tapering off at about 20% of its peak in 2012-2013. In academia it appears to be the

FIGURE 1: Serious games growth in the research field based on surveyed papers in ACM digital library and IEEE Xplore.

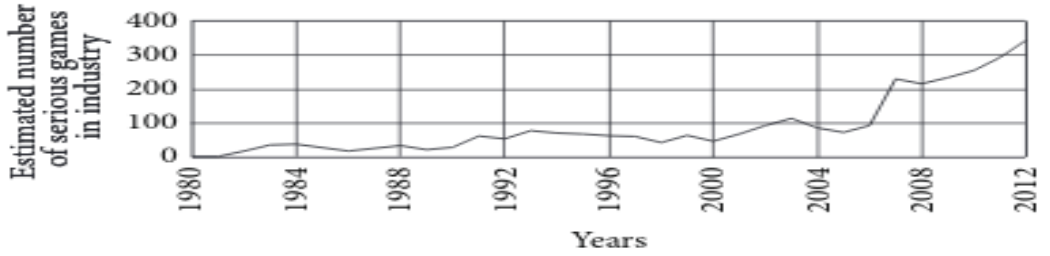


FIGURE 2: Serious games growth in industry.

Figure 3.1 The rates of growth of Serious games overtime with relation to research (Top) and Industry (Bottom) from [1]

opposite, where searching the term 'serious games' in academic databases results in upwards of 50,000 articles on different serious games in the last ten years alone.

Serious games serve a helpful role and can allow for alternative and potentially better insights into problems that do not otherwise have good solutions. There are many different categories of serious games including educational games, games for advertising,

games for physical health, and games for mental health. Some examples are Lumosity™ [2] for brain games, Ring fit Adventure™ [3] for physical fitness, Oregon trail™ [4] designed to teach children specific information or skills, games like Pepsi man™ [5] or Burger Kings Sneak King™ [6] designed to push products, and games like WarCAT designed to help detect for cognitive impairments.

Serious games for education

Serious games used in education coincides with the large number of children devoting major time to games and the potential to use that time to teach children whilst making the education more fun and dynamic, and also reaching a larger audience with remote / online technologies. However, in the past, these games have received a very polarized reception. Some games like "Mario is Missing™" by The Software Toolworks [7], has the objective to teach geography and history to children, using popular game characters, have received very negative reviews, whereas games like Oregon Trail™ developed by the Minnesota Educational Computing Consortium [4] whose goal is to teach American children about the early days of colonization and how to manage resources are relatively popular and receive favorable reviews. Serious games for education are very broad with the literature showing some games being built and designed for classroom use, whereas others were designed to get children invested in learning in their spare time.

There are also serious games for adult education as well with some common examples being for military training and recruitment [1]. Overall, despite early polarization of opinion, serious games in Education are becoming increasingly well received and "serious games proved effective in enhancing cognitive abilities and affect, as well as pleasant mood, in general learning. Through analyzing 46 empirical studies, serious games were reported helpful for learners to obtain cognitive abilities and increase positive affect of learning" [9]. In general, serious games may help with education and increase general knowledge.

Serious Games for Physical Health

Physical health games have existed since the invention of games through sports. However, the topic of serious games for physical well-being is mostly looked at through the lens of physical video games aimed at those who live a sedentary lifestyle or games that utilize a myriad of peripherals to aid in the rehabilitation process. The divide between the games is even more distinct along the research versus industry axis where industry favors the games for promoting good physical health and research seems to favor games

for aiding rehabilitation. Games like Wii Fit [8] and Ring Fit Adventure [3] gamify exercising to get players excited about getting in shape to improve the exercising experience for those that would otherwise be hard pressed to exercise. In a study of the IGER game engine, [10], the purpose of IGER is to allow patients to go through rehabilitation in the comfort of their own homes [10]. IGER uses a combination of many different gaming peripherals like the Wii Balance board and the XBOX Kinect to create a game that makes the rehabilitation process fun and engaging while making sure that the patient is doing the correct exercises properly. Another study details a way serious games can be used in combination with the XBOX Kinect to aid rehabilitation [11].

Serious games for Mental Health

Serious games for mental health are another large area of study focusing on several different and distinct angles: *maintaining* good mental health, *diagnosis and prediction* of mental health issues, *prevention* of cognitive decline, and *rehabilitation* of mental health concerns. One study developed a novel serious game aimed at detecting Alzheimer's Disease (AD) [12]. The researchers utilized the game to target for the common symptoms that can be found in AD, specifically memory and spatial awareness. The tasks involved patients navigating virtual worlds and completing a basic checklist of items that are common for a regular person to do in an ordinary day. The tasks were to make breakfast, then navigate to the grocery store to do some shopping with a list, then navigate back home and do some gardening. The study showed a significant difference between the control group who didn't have AD and the group with confirmed AD. The major difference was in the time it took to complete each task where those with AD took longer to process and recall information. Another study examined a collection of serious games for assessment, rehabilitation, and prediction of dementia [13]. The research looked at assessment utilizing standard psychological testing using the Montreal Cognitive Assessment (MoCA) and the Minimal Mental State Exam (MMSE) and compared the results to what was found in their serious games. What they found was that MoCA and MMSE got the same results as several the serious games presented but in a much shorter time frame [13]. However, it can be seen in several of the games reviewed in [13] that it may be possible to detect MCI.

More of the games in the study looked at rehabilitation of MCI and dementia [13]. The first of the rehabilitation games test a user's visuospatial comprehension using obscured objects, and tests spatial memory using a simple card memory matching game, and finally tests attention using Schulte's tables [13]. Another study used a mobile game and physical Near Field Communication (NFC) with a playing board where participants need to move the phone to the correct spot [13]. However, the common thread in all the

studies is that they were unable to predict cognitive decline and were unable to permanently prevent cognitive decline.

Studies were also reviewed that explored games designed to predict the onset of dementia prior to diagnosis. One serious game employed a word matching game where a collection of letters was to be turned into as many words as possible [13]. Reviewers found that "verbal fluency itself represents only a small part of the neuropsychological assessment spectrum and in itself does not provide any conclusive or comprehensive diagnosis of dementia, particularly not its prediction." [13].

These studies are interesting as they reflect the assortment of paid apps in online stores that claim to have the same objectives in supporting cognitive health, and at times, to assess one's cognitive ability in relation to one's peers. The best or most popular mobile apps at the moment for cognitive training appear to be either Lumosity™ created by Lumos labs [2] or Elevate™ created by Elevate labs [14].

The serious game developed and investigated in this thesis is denoted WarCAT, designed to explore the potential to detect MCIs in the average person while still attempting to maintain the needed amount of entertainment so that players want to continue playing it. It is recognized that this is a very ambitious undertaking, and the efforts made here contributed to identifying the significant challenges serious games face as a means to help assess cognitive decline. The benefits of a game-based assessment approach are that they can be deployed to the general public via the use of a mobile application, which would then allow for gathering data on a larger scale. Data on a large scale are required, as fundamentally these types of approaches rely on statistical inference. In addition, getting the app to smart phones and tablets would allow users to help assess or test themselves for mild cognitive impairments symptoms, changes over time, and provide triggers for clinical assessment. The other major reason for using a smartphone or tablet as the platform for the game is that, mobile gaming is by far the most popular form of gaming at the moment, boasting numbers around 3.9 billion people with smartphones and around 82.98 billion total downloads for mobile games grossing $116 billion US dollars [15].

References

[1] Laamarti, Fedwa, Mohamad Eid, and Abdulmotaleb El Saddik. 2014. "An Overview of Serious Games." *International Journal of Computer Games Technology* 2014 (October): 15.

[2] "Lumosity." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Lumosity.

[3] "Ring Fit Adventure." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Ring_Fit_Adventure.

[4] "The Oregon Trail (1985 video game)." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/The_Oregon_Trail_(1985_video_game).

[5] "Pepsiman (video game)." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Pepsiman_(video_game).

[6] "Sneak King." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Sneak_King.

[7] "Mario Is Missing!" n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Mario_Is_Missing!

[8] "Wii Fit." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Wii_Fit.

[9] Zhonggen, Yu. 2019. "A Meta-Analysis of Use of Serious Games in Education over a Decade." Edited by Michael J. Katchabaw. International Journal of Computer Games Technology 2019 (February): 8. https://www.hindawi.com/journals/ijcgt/2019/4797032/.

[10] Pirovano, Michele, Pier Luca Lanzi, Renato Mainetti, and Nunzio A. Borghese. 2013. "IGER: A Game Engine Specifically Tailored for Rehabilitation." In Games for Health, 85-98. Wiesbaden: Springer Verlag.

[11] Muñoz, J., J. F. López, and J. F. Vilada. 2013. "BKI: Brain Kinect Interface, a new hybrid BCI for rehabilitation." In Games for Health, 233-245. Wiesbaden: Springer Verlag.

[12] Vallejo, Vanessa, Patric Wyss, Luca Rampa, Andrei V. Mitache, René M. Müri, Urs P. Mosimann, and Tobias Nef. 2017. "Evaluation of a novel Serious Game based assessment tool for patients with Alzheimer's disease." PLoS ONE 12, no. 5 (May): 14.

[13] Kazmi, Sayed, Hassan Ugail, Valerie Lesk, and Ian Palmer. 2014. "Interactive Digital Serious Games for the Assessment, Rehabilitation, and Prediction of Dementia." International Journal of Computer Games Technology 2014 (November): 11.

[14] "Elevate (video game)." n.d. Wikipedia. Accessed March 6, 2022. https://en.wikipedia.org/wiki/Elevate_(video_game).

[15] Knezovic, Andrea. n.d. “147 Mobile Gaming Statistics for 2022 That Will Blow Your Mind.” Udonis. Accessed February 15, 2022. https://www.blog.udonis.co/mobile-marketing/mobile-games/mobile-gaming-statistics.

Chapter 4: Game Development

Card Game

A card game was selected as the basis for this work as card games are universal and typically very easy to learn and play. The idea behind this card game in particular was to base it on the very simple and likely familiar card game War [1]. The standard game of War is a game where a deck of cards is divided between two players. Each player draws a card from the top of their deck and immediately plays them against each other. The higher card wins, and in the case of a draw, one keeps drawing from one's deck until a winner is determined.

In the game designed and developed here, we take that simple game and make it slightly more complicated to allow for strategy to be employed, but still simple enough that anyone can learn it. Instead of dividing the deck into two, each player is dealt a hand of five cards from the same deck. Players can play the cards in their hand in any order but can only play each card once. Once both players have selected the card they wish to play for the current round of battle, the cards will be compared against each other, and the higher card will win.

In this game, a human player is playing against the computer (bot). Each player is dealt five cards. When each player plays one card, it is called a *round.* A *game* consists of five rounds. A *level* consists of either 50 or 100 games depending on the version.

The game's scoring has also been altered to favour a narrow win over a more decisive win. After a card is played and compared to the opponent's card, the larger the disparity between the cards results in fewer points to the winner. For example, if the player played a King against a 2, the winner would only get two points, whereas if the player played a 3 against the 2 the winner would get 12 points. The equation for score calculation is

13-(highcard - lowcard)

where Ace = 14, King = 13, Queen = 12, Jack = 11, and number cards equal their face value.

This scoring calculation leads to the idea that you have to play risky and anticipate what your opponent will play and know when to sacrifice points where necessary in order to develop a winning strategy. These simple changes change the nature of the game

significantly and allow for strategic play. By allowing for more strategic play, it makes it clearer who is learning a strategy and who is playing in a more random-like fashion.

There are a number of different play modes to complement the game. The one that was primarily used and tested was the single strategy tournament version, where the randomness has been removed and the computer BOT has set hands and a set order they play the cards (e.g. from the highest value card to lowest value card). Players competed in the game by trying to get as high a score as possible in a limited number of games. Figure 4.1 shows the information screen that new players are given to read on starting the game.

The next section will describe some of the iterations and successive refinements that were undertaken during the design development of the WARCAT game. Its genesis was from a game previously developed to assist in assessing gambling addictions, modified here to explore its potential to detect and assess MCI.

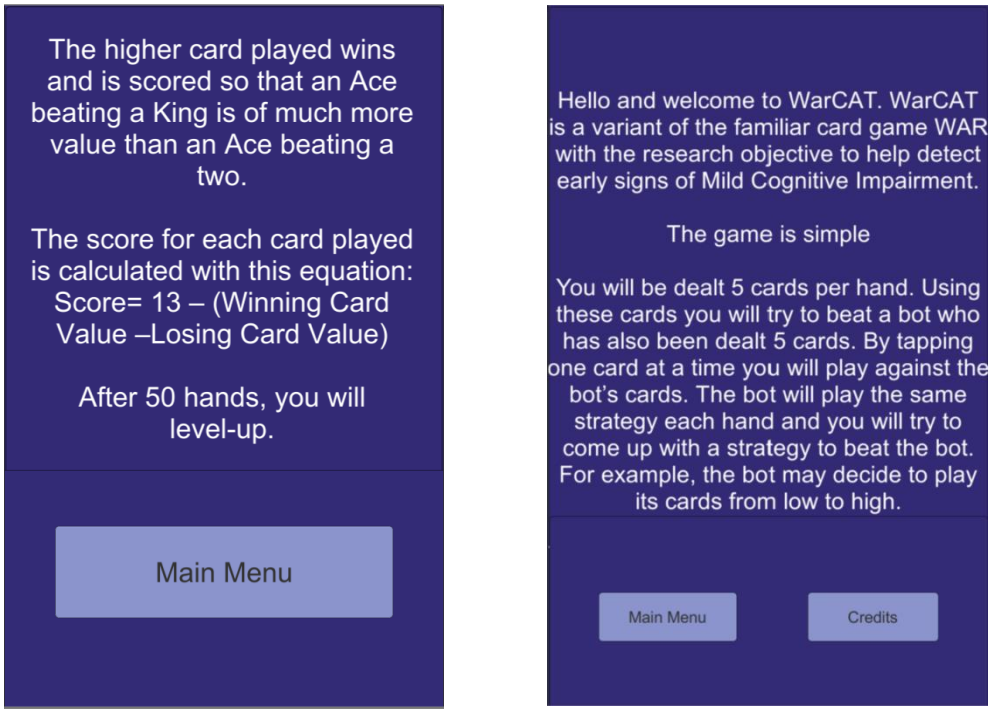


Figure 4.1 The information screen players see when they start the game

Play styles

WarDOG version

Before becoming a game aimed at detecting or assessing MCIs, prior work focused on the potential for serious mobile games to identify gambling addiction, using a prototype game developed called WarDOG (War Data on Gambling). WarDOG differed from WarCAT in a few key ways. The major change was that in WarDOG, a player earned in-game currency that was designed to simulate the feeling of winning at gambling. This was meant to mimic the 'freemium' style of gameplay that is ever present in mobile games today, where players can effectively pay to win or in this case pay to get prizes. In WarDOG, freemium style was mimicked to see if it elicited addictive behaviors and tendencies in players. This in itself is a very ambitious undertaking, but it is well established that many games elicit addictive behavior although reasons for this are speculative at best.

The largest difference between WarDOG and WarCAT was that there were a few exclusive modes of play to WarDOG. The exclusive modes of play were a Player vs Player (PvP) mode where players would directly compete against each other, and a tournament mode where multiple people would get slotted into a bracketed in a single elimination style tournament. Both modes of play were designed with the feeling of gambling and addictions in mind and are less useful in detecting MCIs. WarDOG also makes use of the various sensors and other mobile game tricks to enhance addictiveness. It is from this development and initial experimentation that a serious game based on a simple card game was hypothesized to have the potential to detect MCI, as it involved recognition, memory, strategy development and related cognitive functions.

WarCAT version

War Cognitive assessment tool (WarCAT) is the card game designed to explore its potential to detect MCIs, rather than for the study of gambling addiction like WarDOG. A key difference in WarCAT is the absence of direct player vs player modes that exist in WarDOG. WarCAT utilizes the idea that emergent strategy and memory patterns could be collected through game play, and that these data may shed light on cognitive function and MCIs. WarCAT has players playing against simple BOTs that only utilize one strategy for a large number of games. After players are proficient enough at the game to win consistently, the player will level up and the BOT would switch its strategy. More specifically, the game play data are collected with the future objective to compare and classify them for their potential to detect a player recognizing that the BOT is playing a

set strategy, recognize the strategy itself, adapt one's own play (strategize) to counter the bot's strategy, and remember this learned strategy. This classification is beyond the scope of this thesis; however, the thesis developed the game in a manner that the game play data are collected and available.

The assumption is a typical player should experience a winning percentage of approximately 66% if they can create even a simple winning strategy. We conjecture that a player with MCI would play below this threshold. If play is statistically different enough from the norm, it may indicate that the person is experiencing some degree of cognitive decline, exemplified in the game play data as a random strategy as opposed to developing and remembering a more effective winning strategy. Of course, there are exceptions to this, as a major component of the game is that the cards dealt are random. However, this factor can be controlled to a degree by either making every player experience the exact same games or by using the law of large numbers so that the randomness can be overcome (the variance diminishes over a large number of games). WarCAT explored both methods to handle the randomness problem.

WarCAT records information on when people won when they should have won, lost when they should have lost, won when they should have lost, and lost when they should have won. This information is recorded by considering what everyone's hands are and what all the possible combinations of plays are to get the exact odds of winning any particular hand. This information in combination with the large number of games we expected people to play, would provide the data needed to compare an individual's play to others in the same demographic. Another option of controlling the randomness is explored in a version of WarCAT called WarCAT Single Strat. WarCAT Single Strat works exactly as its name implies; players have set randomness and the BOT opponent only ever employs one strategy for the duration of play.

Game mechanics

Micro transactions and points

In WarDOG, an in-game currency was needed so that the true environment and feel of gambling could be recreated in a mobile game format. The currency system is an in-game points-based currency that is won through playing the game naturally. The points are directly tied to your score, so the better one plays, the more points one could earn. On the other hand, if the player loses, they will lose points as well, which can lead to debt in the game. In the various forms of play, betting was added so that when the game

is played against a real person, a player can bet on their skill to earn more currency. The tournament mode of WarDOG also had an upfront fee to enter, and only the winner could reclaim their fee in addition to the pot. The matchmaking for the PvP was random as well, so unless there were only two players in the world playing, you could not use the system to gain infinite amounts of points, at least not without robbing someone else.

The WarDOG continued with the idea that people could buy the points with real money. The reason for this is to emulate the popular format of mobile games at the moment typically referred to as freemium, where one can play the base game for free, but if the player wants to speed up the game or enhance their enjoyment with additional features or functionalities, they can pay money. To further emulate modern mobile games, microtransactions were added so that players could pay to change up their cards to show off to others. Card packs were also going to be added to give people the rush that gamblers would get from a good round of gambling, and to better look at whether this style of giving rewards truly counted as gambling or not. The emulation of the mobile gambling experience extended to a lockout mechanic where just by playing the game naturally, a player would be locked out for a set period if the player did not spend money. Figure 4.2 shows the overlay players see when they are locked out. In WarDOG it was set up so that if the player lost a game, they would lose points proportional to the loss, and if they lost enough points they would be locked out of the game, slowly recuperating points at a rate of one point per minute with a maximum lockout time of 24 hours, after which they would be reset to zero and allowed to play the game again.

The other option a player had was to either pay money to the game so that they could recoup those lost points immediately and begin playing again, or they could complete a simple survey on addictive tendencies and preferences. The surveys were created and integrated into the game as a way to assess whether player's beliefs and descriptions of their own playing habits matched what the game-play data revealed about their playing habits. The survey questions were created with input from sociologists, and the survey's inclusion in the game was reviewed and approved by the Research Ethics Board.

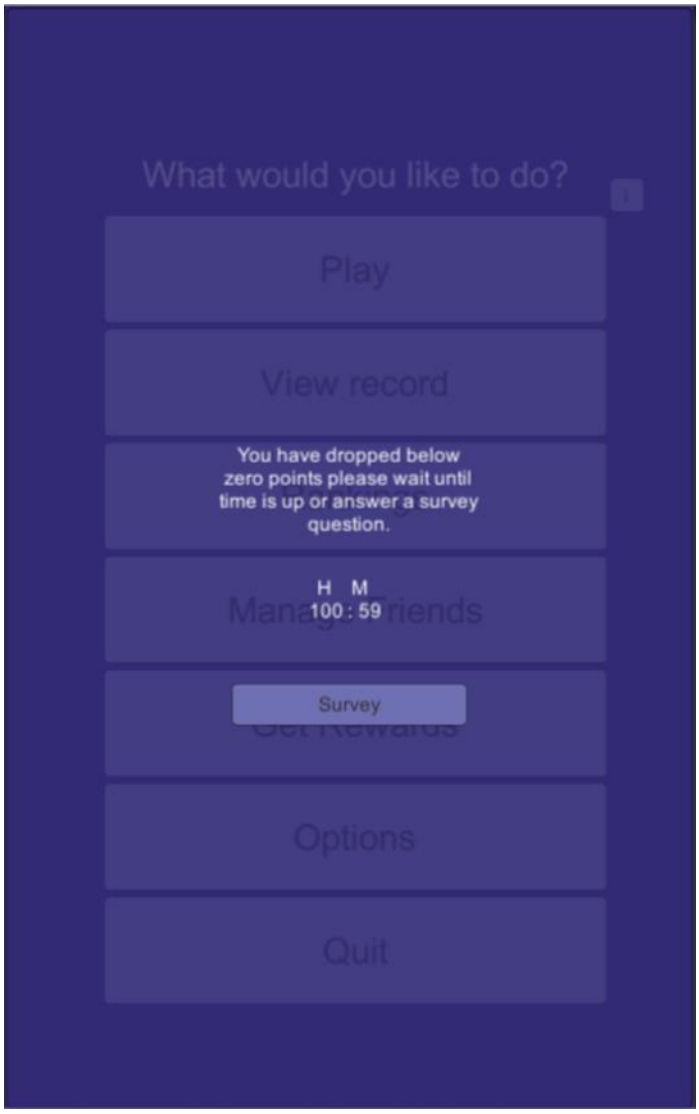


Figure 4.2 The overlay players see when they are locked out.

The actual payment method was never finished so the only option available to players was to wait or answer a survey question. This lockout and payment were done to simulate the situation where players could sit through advertisements as opposed to opting in for a small fee. It became quite evident that the main feature of any game is its addictiveness which engages a player to continue playing. In hindsight, this goal to create an engaging game with the worthy goal of assessing addictive behavior was unrealistic in the limited resources of the research group as opposed to the resources available to gaming companies.

Reward points

The in-game currency that was used in WarCAT and WarDOG are referred to as reward points. In WarCAT, the reward points were used to keep track of how well a player is doing in any given match, as the more reward points you earned, the better you played against your opponent in the weighted scoring described earlier. In WarDOG, the reward points were used as an in-game currency as described earlier. In WarCAT, using reward points in addition to simply counting wins and losses, allow an inference of how well someone is learning the game. What we expected to see from the reward point system is that a player's initial wins and losses would be at about a 50-50 rate. This win rate should improve as they recognize the bot's strategy and learn the basic countering strategy, and one should also see an increase in the amount of reward points earned in each round of play. The maximum number of points per round of play is 60 and the average number of points for well thought-out strategic play is closer to 20-30 points. Random play should lead to approximately zero points if we include the loss of points from losing a hand.

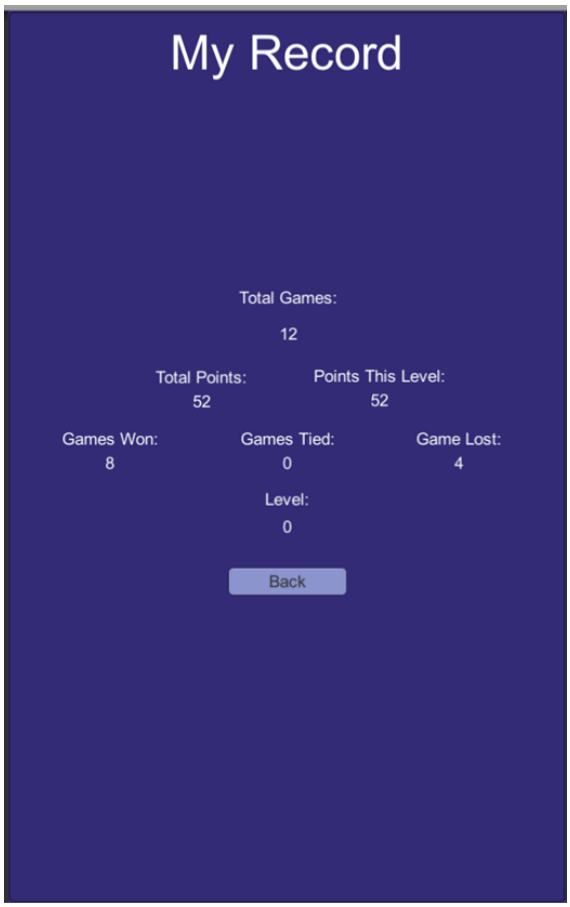


Figure 4.3 The record panel detailing all the information a user gets.

Figure 4.3 shows the record panel detailing all the information presented to a user on their progress. The amount of reward points earned after each game is calculated by the difference between the player's score and the bot's score to promote high risk strategies even more. To see whether players are playing to the best of their ability, we calculate the total possible amount of reward points they can win as well as how many possible winning hands there are, and then compare those numbers with what they actually achieved. Keeping track of the number of winning hands allows us to see how much skill they have developed. However due to the randomness inherent in the hands dealt, the odds are almost impossible that any player can get a perfect score. This results in each game having a unique maximum amount of reward points and score that are possible to earn. These metrics, in combination with others gathered, allow us to extract an inference on a player's cognitive level with respect to game play. It also allowed us to track certain game play methods discussed in more detail below, such as the race mode and the single strategy tournament that was run for testing purposes. We used the reward points in both cases to see who was performing at a higher level of play and strategy. Getting this information was extremely useful in determining the optimal methods of play and whether they were different than the expected optimal strategies.

Level progression

Another major component of WarCAT is the leveling system. The idea behind the leveling component is that in order to progress to more difficult play styles, players have to play better than 1 standard deviation of winning play. In simpler terms, players have to win 68% of the matches in a particular playstyle to be moved to a harder playstyle. This leveling system allows us to see where and how people are struggling. The idea would be that at each level, a learning curve should be seen as players identify and learn a strategy that will win at each level. This level system, in combination with the reward points and overall game play, allow an inference on someone's level of playing ability and memory. If they have good mental abilities, they should be able to clear each level faster and faster. In the first level, the simplest and easiest-to-identify BOT strategy was employed. The BOT would simply play its cards from high to low. Unless players identified and countered this strategy, they would likely lose as often as win by playing randomly. Playing randomly against the BOT would result in a score comparable to the BOT's, which would cancel each other out. Table 4.1 shows players playing through the different levels and their numbers of wins and losses at each level as well as the percentage of wins to losses that they achieved when they levelled up.

User	Level 1		Level 2		Level 3		Level 4	
	win	loss	win	loss	win	loss	Win	loss
Participant 1	142	53	220	74	292	101		
	72%	28%	74%	26%	74%	26%		
Participant 2	70	30						
	70%	30%						
Participant 3	85	12	160	34	219	50		
	87%	13%	84%	16%	81%	19%		
Participant 4	80	16					317	72
	83%	17%					81%	19%
Participant 5	69	28	208	87				
	71%	29%	70%	30%				
Participant 6	80	19	162	35	256	51		
	81%	19%	82%	18%	83%	17%		
Participant 7	75	26						
	75%	25%						
Participant 8	76	23	149	50	213	83		
	76%	24%	74%	26%	72%	28%		

Table 4. 1 Players wins and losses after completing each level of the game

The second level had the BOT playing its cards from low to high and was designed to throw players off momentarily as they have to identify and counter a new strategy. The final strategy had the BOT playing its cards from the middle value and going down, finishing with its high cards. This made for a more difficult challenge to play against. If one played to sacrifice a card at the start to gauge the BOT'S hand, one would likely be giving the BOT a lot of points, making it hard to win from that position. Conversely, if one started with their highest card, there is the chance that they would not earn a large number of points. In between the different levels, there is a congratulations screen where we tried to show players what we thought their strategy was by mapping how often they

played a particular card in their hand each round. This was represented in a bar graph of the cards in hand organized from high to low on the y-axis and on the x-axis was when the card was more likely to be played.

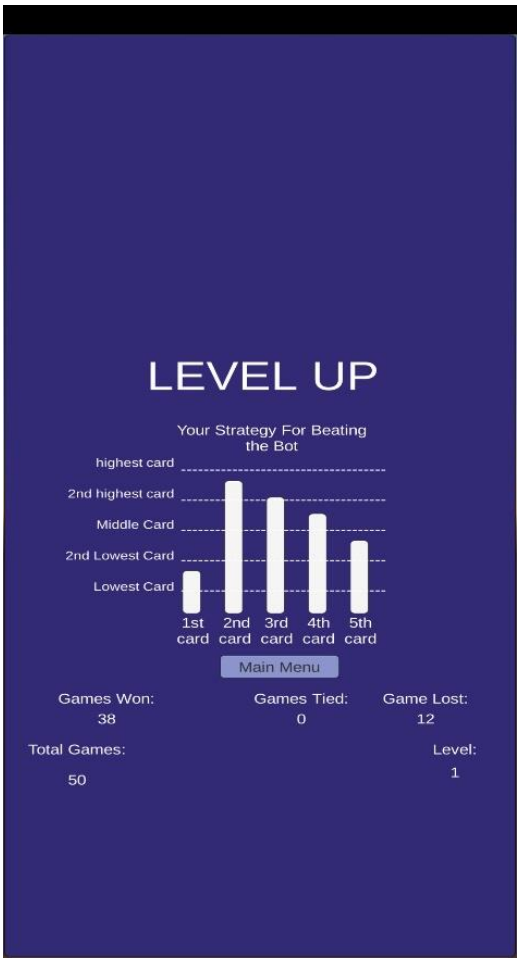


Figure 4.4 The level up screen

Figure 4.4 shows the level up screen and an example of a strategy. The main purpose of the graph was to graphically infer whether a player was developing a strategy or if they were just playing randomly. This idea was good in theory; however, it was executed poorly as in all the real-world tests, the graph tended to show only randomness, as people tended to adapt their strategy for their hand rather than just figuring out a simple overall strategy of card values in relation to each other. Later, a line graph was added that mapped scores over time so that players could see their progression. The idea of showing players their progression was adapted and turned into a graphical fingerprint. This fingerprint would take in their play data and determine whether the strategy was random, simple, or advanced. It then took all the game's strategy values and placed them

into a circle pattern making it easy to identify likely instances of forgetfulness, and since it was now an image, later on it will be easier to classify with machine learning methods if desired.

Game Modes

Player Versus Player (PvP)

The Player versus Player (PvP) mode is exclusive to WarDOG. Figure 4.5 shows the game play panel for PvP without anything loaded. The PvP mode is set up so that two players on different devices can compete against each other for increased in-game currency. In the PvP mode, players are given a number of options including wagering and chatting. Wagering is used to increase the amount of your reward points by betting on yourself. The idea behind having wagering is to mimic gambling games to better study gambling behaviors and addictions. The chatting is meant to simulate a poker-esque environment so that bluffing and minor mind games can be used. The chat feature is limited to a few simple buttons that send a random preset message to the other player to make sure nothing too uncouth is said, these messages include good game messages, good luck messages, bluffing messages, and taunting messages. In addition, to promote spending the in-game currency, players could pay to cheat by peeking at their opponent's hand. PvP could be implemented in WarCAT, however it would be hard to interpret results from this type of play and the features don't necessarily add to the overall objective of assessing potential to detect MCI. The reasoning is that if the player plays against a set opponent where it is possible to know exactly how it will play, their level of cognitive ability can be gleaned. While PvP would allow people to develop unique strategies, the amount of randomness and the lack of information makes it more akin to gambling rather than formulating easy-to-distinguish strategies.

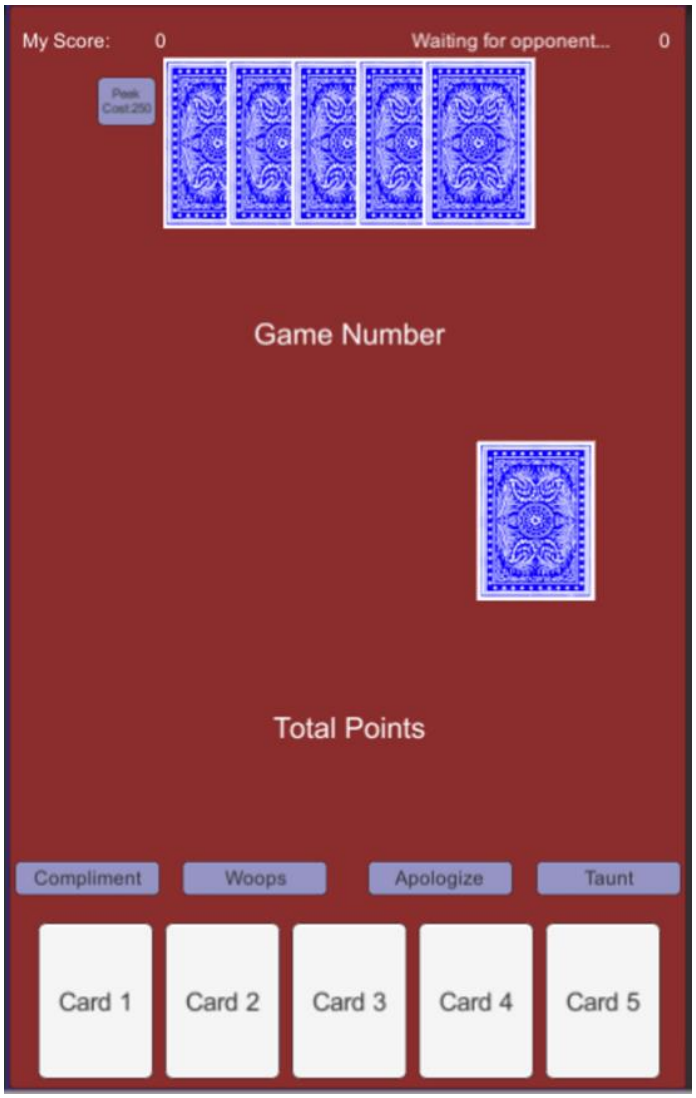


Figure 4.5 The gameplay panel for PvP without anything loaded

In the PvP mode, the servers are set up so that computation happens on the server end, and the results are sent back to each players' respective device so that there is very little lag. The other reason everything was handled by the server was so that all information can be properly organized and stored faster. The server also handles the chatting between the players, and it keeps track of how much each player has bet so that the proper amount of reward points can be allocated. There is also a play timer on how long players have until they are forced to play a card so that they cannot hold up their opponent and so that matches are relatively fast. The maximum time players were allotted to play a card was 15 seconds. The fast play is designed so that players can get into a game and out of a game in a relatively fast manner so that they want to just keep playing the game, similar to how people will just keep playing slots due to its mindlessness, speed, and short-duration. There were plans to have a system that would differentiate different

levels of play ability so that players only play with people of their own level; however, that was never implemented before WarDOG was replaced in favor of further investigating WarCAT.

Player Versus Bot (PvB)

This game mode exists in WarDOG and in WarCAT; however, in WarCAT the PvB is slightly different. Figure 4.6 shows the PvB screen in play. In WarDOG, the PvB mode was a filler mode so that people could play offline and by themselves. This mode had all of the same features as the PvP gambling mode with the exception of the chat and wagering features as those were unnecessary to implement.

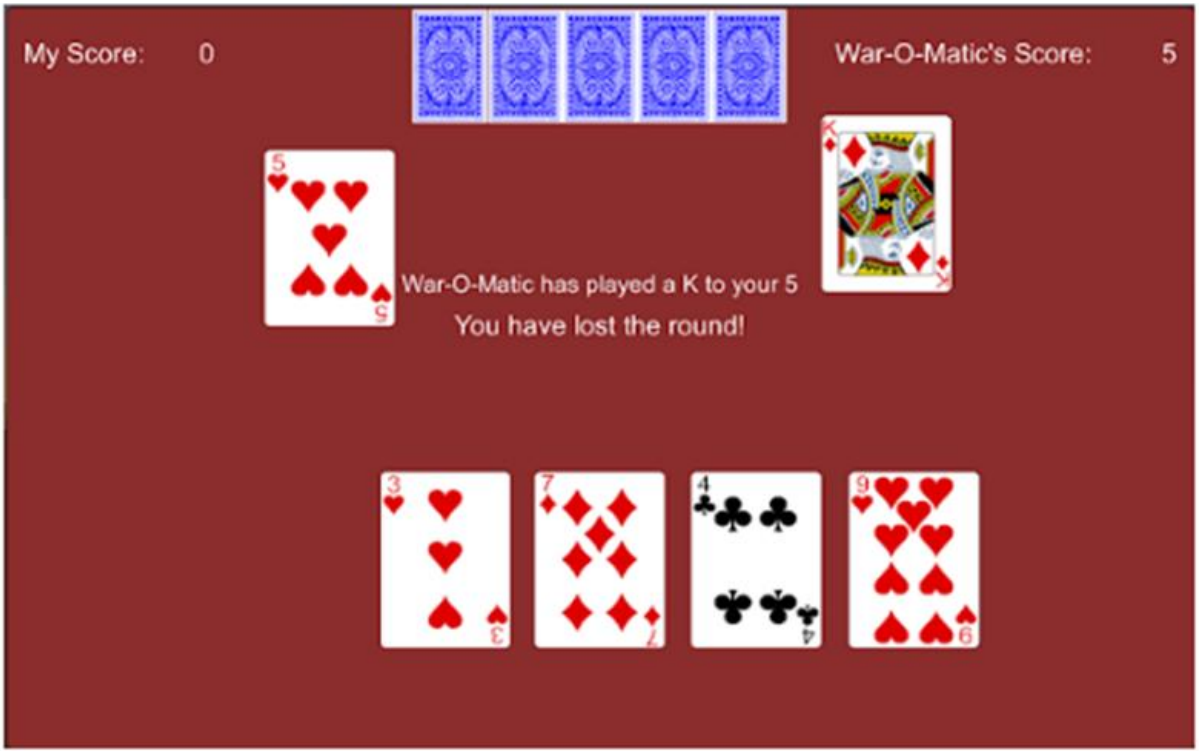


Figure 4.6 The PvB screen in play

For WarCAT, the PvB mode was far more important as there was no PvP in the game at all. In WarCAT, all of the gambling functions were completely removed, however the gambling currency was changed to reward points as stated earlier and more styles of play were added. Different levels of play were also added in here. PvB serves a crucial purpose in WarCAT: it allows us to track how players are doing compared to one another without having to rank them for competitive play, and it allows us to track how players are doing when compared to how they did previously against the same stable opponent. The biggest benefit of the decision to only have a PvB mode is that the games become extremely fast so collecting a lot of game data becomes a lot more feasible. The PvB mode

is separated into the 3 levels mentioned earlier. In order to progress a level, the player must play at least 100 games and must win at least 60 of them. The win rate is to show that the player can develop and then play a strategy sufficient to get significantly above the 50% win rate that random play would get. For WarCAT, the server was rebuilt, and everything was put into the actual game application so as to remove as much communication between the server and mobile device as possible. Moving all the game features allowed for the future implementation of an offline play mode where data were stored locally until such a time as it could be sent to the server. It also allowed for the ability to save login information right on the phone directly, meaning the user would only ever have to log in once.

Tournament

Another WarDOG exclusive mode was the tournament mode. This mode functions almost exactly the same as the PvP with the exception that players need an entrance fee to participate. The entrance fee was then used as a pot for the winner of the tournament. The tournament was a bracketed one-elimination style tournament. The point of the tournament was to get people gambling on themselves to better simulate a tournament gambling environment. Furthermore, we also ran a tournament where people were competing to get the highest score and win rate possible in the smallest number of games. This tournament served to gather data on what the most effective tactics to beat the game are and to see whether that data matched our assumption that playing a sacrificial card to discern the opponent's strategy was the optimal strategy.

What we learned was that a single sacrificial card was not always optimal, and that players needed a more adaptive strategy and a bit of luck to get the highest possible win rate and points. The backend of the tournament mode for WarDOG was a complicated timing system where the first player in a tournament would create a queue and then players would enter into that queue later until it was either full or a tournament start timer would end. After the timer ends, any empty spots left would be filled with bots of random skill levels. In the case of someone dropping out of the tournament, they would lose their entrance fee and their slot would be filled with a bot in such a way that the tournament wouldn't notice with the exception of the person playing them. In the case of everyone dropping the tournament, the queue would be cleared and destroyed. Figure 4.7 shows a tournament game in progress.

Figure 4.7 The Tournament screen in progress

Single Strategy

The single strategy mode is exclusive to WarCAT. As the name implies it is a simplified version of the game where the bot only plays one strategy (e.g. high-to-low, low-to-high, or similar). This mode is for testing purposes and for gathering data that could be more easily applied to machine learning as it removes some of the variability of

the system. When distributed to the testers, it allowed us to see whether our speculations on most effective strategies and most effective game-play were supported. This version was played in a large tournament setting where numerous testers were pitted against each other and the winner of the tournament would get a larger sum of money than the other competing testers. It was used in conjunction with a version of WarCAT where all variability was removed, meaning all testers would receive the same opposing hands to beat, thus giving us the best idea of how standard strategies, advanced strategies, and novice play looks. To make this mode work, the number of games played was used to move the random seed forward appropriately so that players couldn't quit out of the application and play the same round over and over again by resetting the seed. For the single strategy, the bot would only play from high-to-low which was deemed as the easiest to learn playstyle. What we found was that our expectation on optimal play was entirely unsupported.

The basic assumption was that to get the most points consistently when the opponent is playing high-to-low, you should sacrifice your lowest card to start and then proceed to play highest to lowest. This method was assumed to be the best as it could be seen that if one employed the strategy, one would give up a small number of points to the opponent in exchange for being able to consistently play a card above the opponent's highest card, maximizing one's high value cards. What was witnessed, however, was a far more adaptive strategy where depending on what cards were in the player's hand, they would either play a high card first to test the opponent's hand, then high to low, or they would play more than one sacrificial card. With the help of Dr. Kai Gutenschwager, we were able to map different styles of play with different levels of strategy to determine the level of play, and we then monitored the different adaptive strategies the player used over the course of 100 games. This gave us a far better fingerprint with which to train classifiers. This mode also allowed us to see how engaging the PvB play was, in anticipation of the app eventually being distributed out to the public, as well as how fast someone could get through a level.

Race Mode

Race mode was a simple idea for a mode to enhance the longevity of the game, change up the playstyles, and allow us to test the players in a different way. The race mode is played in a similar manner as the PvB mode. In the race mode, one plays to complete the game as fast as possible with the goal being to reach a total 500 reward points. The player is given a five-game grace period where they will not lose any points so that they can get familiar with the game and not suffer from point lockout. In the same way as a time trial in racing games, you do not play against an actual person but rather compete

against one's own records. This race mode allows us to parse data more easily as the data gathered are automatically ordered into a highest to lowest time order.

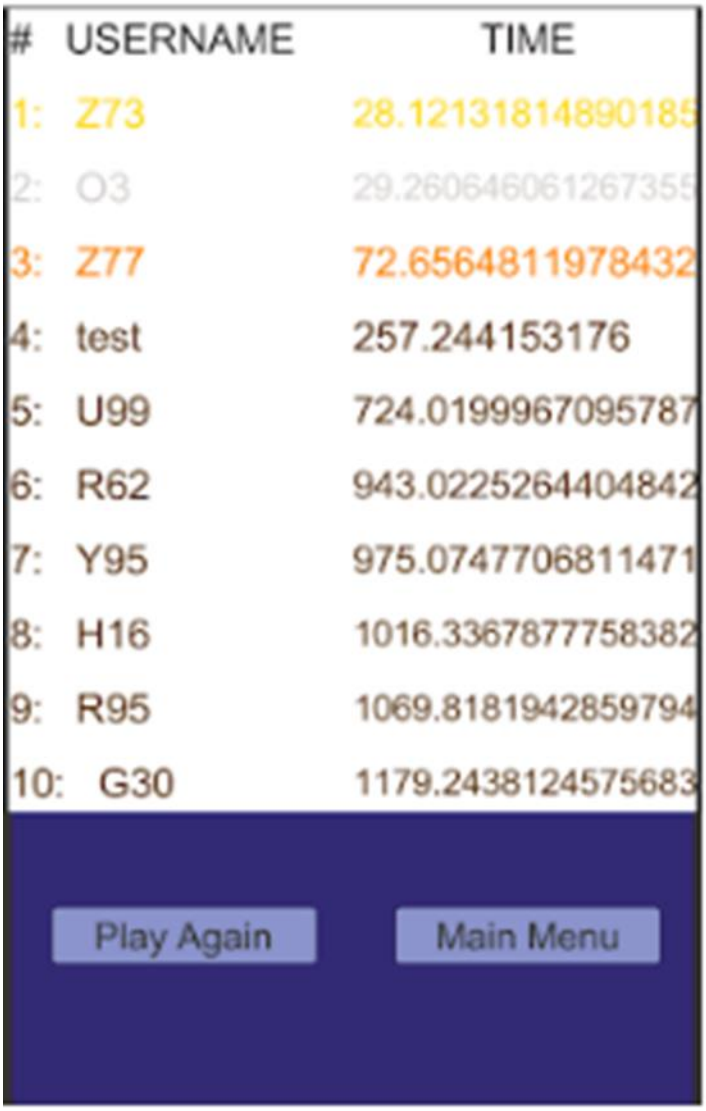


Figure 4.8 The leaderboard for racing organized by how long it took to complete the game

Figure 4.8 shows the leaderboard for racing organized by how long it took to complete the game. It should be noted that the leaderboard fills its table with fake accounts that represent variable scores with first and second place holding the best average times based on data gathered from the games in PvB mode in the database. It also allows for easier comparison for the purposes of machine learning because it is very clear who has developed a winning strategy, who is playing randomly, or who is having trouble with fast processing. Currently people don't race against a specific person of choice but

instead pick a BOT strategy to play against. A future evolution for this mode is to lock the variability in the same way as with the Single Strategy mode, and add a feature where people can either enter in a specific seed for the deck and compete in that seed, or pick a person to compete against directly.

The other major objective of the race mode is to put people into a higher stress situation where they actually have to play at a higher level very quickly. Forcing players to play at a higher level is assumed to stress cognitive function making them rely on creating simpler strategies, and focusing more on quick decision-making skills. Keeping track of the number of winning hands also allows us to see how much skill a given player has at the game. This is mainly due to the nature of how cards are dealt and the preset maximum amount of points a player can earn during the game. If we compare what they earned to that number, we can get a rough idea of their skill level in comparison to other players. The exact data recorded in this mode is player score, number of games, the number of wins, losses, what strategy the player was racing in, and the total time it took to complete the game. This mode has a visible in-game leaderboard so people can see where they rank, and what scores and times they need to beat to improve. The leaderboard could also be adapted very easily to allow for set seeds and comparisons of certain seeds. Race mode was implemented in WarDOG but was designed to be used to test for both MCI and gambling addiction.

Peripherals

Through discussion with professionals in the field of dementia and Alzheimer's disease, the idea of measuring different biomarkers using peripherals came up. The peripherals that they were using for their studies were things ranging from exercise equipment, to physical sensors and more. While discussing the topic of peripherals, the many symptoms of Alzheimer's Disease came up. Two feasible options for potential peripherals were heart rate and temperature sensors [2]. These two sensors were chosen under the assumption that if one had a mild cognitive impairment, dementia, or something in between, they would likely get flustered while playing the game if they get confused and struggle to remember what they were doing, and we could likely expect to see heightened heart rates and temperatures. This assumption still needs more field testing, however after visiting and communicating with older adults as well as anecdotal encounters with a family member with MCI, I observed inadvertent lying because they either forgot they did something, or they were embarrassed about what they did. The other point of note is that older adults with cognitive decline would get exhausted very quickly from anything that taxed their cognitive function.

The peripherals that were actually tested were put together by Genico Melegrito and programmed by the author. Figure 4.9 shows the heartrate and temperature sensors in use [2], and Figure 4.10 shows the Arduino schematic made by Genico Melegrito [2]

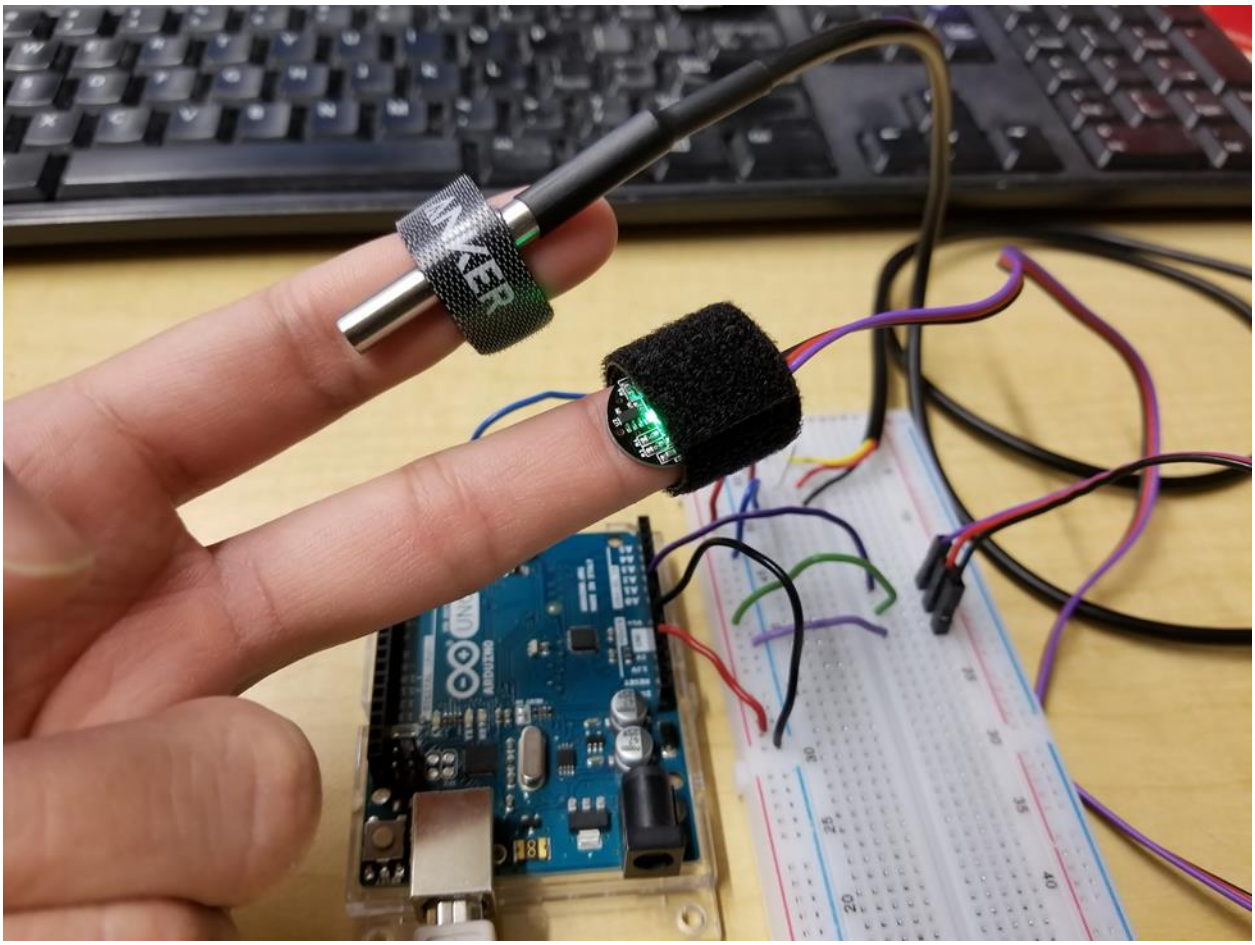


Figure 4.9 Temperature and Pulse sensors used while playing

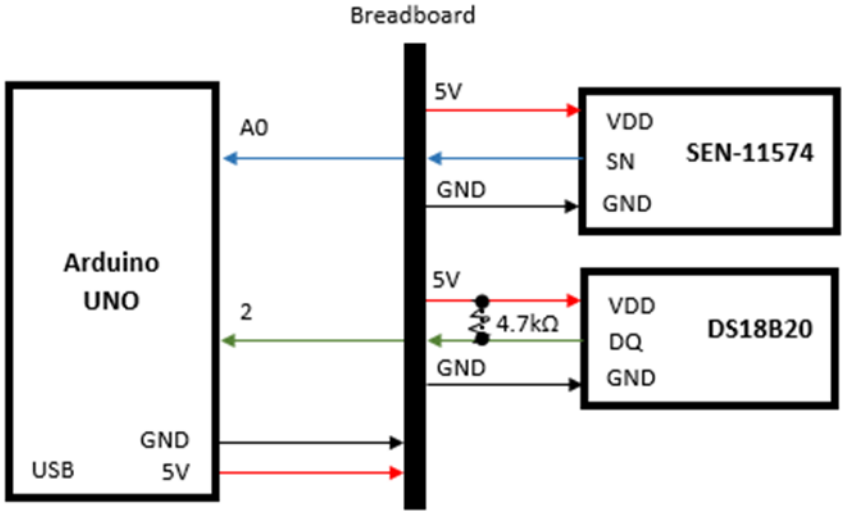


Figure 4.10 Simple connection diagram for the sensors

Heart monitor

The Heart Rate monitor was a simple pulse sensor connected to an Arduino uno and was to be used in conjunction with the temperature sensor to measure player's heart rate and temperature, from which to potentially infer how nervous the player was, and to see whether these data correlated to cognitive abilities reflected in the game play data. If increased stress affected the cognitive abilities, it may mean that there could be false positives in the classification of the data down the road, giving these sensors an invaluable role.

The pulse sensor was a SEN-11574 connected to an entirely separate Arduino UNO board and attached to the player with a simple Velcro strap. The Arduino communicated with a separate dedicated server and database using a similar method as the server for WarCAT. The sensor was set to constantly detect the person's heart rate at a sampling rate of 500Hz and a baud rate of 250000. The Arduino would take snapshots of the person's heart rate throughout the game and then send that snapshot to the server once the game was over. The dedicated server would then submit the code into the appropriate table in the appropriate database. The sensors collect a lot of data in a short amount of time, and our focus was on data collected during actual play.

Temperature sensor

A digital temperature sensor was used to examine its potential correlation with MCIs and gambling addiction. The digital temperature sensor DS18B20 is connected to an Arduino UNO board in the same manner as the pulse sensor. The temperature sensor is meant to be used in conjunction with the pulse sensor. It is connected to the player via a Velcro strap attached to the player's finger with physical wiring to the Arduino UNO. The temperature sensor's main purpose was to see whether potential flutter or excitement was reflected by higher temperature readings, which could add an extra dimension in classification for either MCIs or gambling addiction. The idea as mentioned earlier is that if one has a mild cognitive impairment, they may be more likely to struggle and get embarrassed or flustered, leading to elevated body temperature. The temperature sensor chosen is sensitive enough to distinguish whether the player's body temperature was outside the normal expected temperature of 36-37 Celsius. The temperature sensor, in the same manner as the pulse sensor, was set up so that it was continuously recording but only captured during and after each round of play. The sensor would send the data to the Arduino which was connected to the same dedicated server as the pulse sensor. As stated with the pulse sensor, the server would send the interpreted data to the database for proper storage.

Brain scan device

Another peripheral considered was a device for scanning the brain. The purpose of this device is to scan the brain of a player while they are playing as additional information with no pre-determined purpose, but to be open to serendipitous insights. As the brain scan device cannot be distributed out to app users easily, it would likely need a direct setup which could be monitored by a doctor or clinician making for easier assessments. The issue with this device is that it is clunky and would require a person to play the game in a static setting, which partially defeats the idea of using a mobile application game with wide distribution among the population. The brain scanner would likely be better used with a game designed for a static setting rather than WarCAT. A relatively inexpensive device would be a portable electroencephalogram (EEG). These are relatively simple devices that detect electrical activity in the brain using small, metal discs (electrodes) attached to the scalp.

Eye Tracking

Eye tracking was also considered to infer patients' recollections, using the same tactics employed to detect for lying by police services. The eye tracking would use a mobile device's internal front facing camera to monitor eye movements. This idea was discarded for a number of reasons. The first reason was at the time of discussion the front facing camera was not powerful enough to reliably distinguish the players eye movements. The other major reason was storage space for the files as there would be 100 videos minimum per player each spanning about 1 min in length, and they would need to be in high enough quality to track the movements, for potentially thousands of players. The other option for storage was to store the velocity and direction of movement continually for the entire length of the game. The ongoing evolution in camera technology and data storage makes this idea something to consider again.

Backend

Servers

The server architecture for these games is very simple. The servers are Java based servers developed using the Eclipse IDE. The servers are set up so that they are only interacted with for logging in, logging out and capturing user data. The server receives varying messages, the messages are encoded and decoded upon transfer. The message received always starts with the type of command, followed by the command to be executed and any data to be stored. On the server end, the server has a decoder which then feeds the message into the appropriate message handler. The handler then looks at the command and executes it. The different types of messages it can receive are login requests, registration requests, game logging requests, and log out requests. For the log in requests the server receives username and password information, it then uses the username information to pull up the user from the database and then compares passwords. If the password given is wrong or the user doesn't exist, the server sends back an error response so that the game can let the user know that something is wrong. If the information is correct, it sends an OK message to the game. The game then records the valid username and login so that players don't have to login again (until they log out). The registration works similarly to the login except that it only sends an error back if the username requested already exists, otherwise the server sends back the ok message.

The message containing the game information is more complicated as there are a few different types of messages sent. The first type of game message sent is the user info message, this message stores all of the users' info from their username to what cards they

are playing with. The second message type is the post-game messages which record every detail of every game played by the user, and these data include the strategy the BOT is employing, the cards in each hand, the order in which the cards were played, the odds of winning, whether or not the user won the hand, the amount of points they earned from the hand and finally it is recorded whether or not they used an advanced strategy or not. These data are interpreted by the server and put into the appropriate tables in the database, and depending on the version of the game it will put it into an entirely separate database as well.

The server originally used TCP/IP for communication however on the advice of a lab colleague it was switched over to a more UDP/IP server architecture. With the exception of login and registration where it needs an all-clear receipt to continue, the game cares little for if the server is running or not. There are in total 3 servers running: a server for capturing data and managing login information, a server specifically for peripherals so that the Arduino (if used), and finally, a server for the WarDOG version of the game. The peripherals have their own dedicated server so that the Arduino isn't trying to send data to the server via the mobile device. The WarDOG server handles everything the game does including deciding what cards are dealt, what happens with play and points distribution and everything else that the game application would normally be expected to handle. It might be best to rewrite the servers in C# for better compatibility with the C# used for the actual game; however, there is no noticeable problem with the Java based servers at the moment.

Database

Field	Type	Null	Key	Default	Extra
idGame	int(11)	NO	PRI	NULL	auto_increment
HANDLE	varchar(20)	NO	MUL	NULL	
TIMESTAMP	datetime	NO		NULL	
FIRSTTIME	double(4,2)	NO		NULL	
SECONDTIME	double(4,2)	NO		NULL	
THIRDTIME	double(4,2)	NO		NULL	
FOURTHTIME	double(4,2)	NO		NULL	
FIFTHTIME	double(4,2)	NO		NULL	
FIRST	tinyint(3) unsigned	NO		NULL	
SECOND	tinyint(3) unsigned	NO		NULL	
THIRD	tinyint(3) unsigned	NO		NULL	
FOURTH	tinyint(3) unsigned	NO		NULL	
FIFTH	tinyint(3) unsigned	NO		NULL	
OPPFIRST	tinyint(3) unsigned	NO		NULL	
OPPSECOND	tinyint(3) unsigned	NO		NULL	
OPPTHIRD	tinyint(3) unsigned	NO		NULL	
OPPFOURTH	tinyint(3) unsigned	NO		NULL	
OPPFIFTH	tinyint(3) unsigned	NO		NULL	
STRATEGY	tinyint(3) unsigned	NO		NULL	
WINPERCENT	tinyint(3) unsigned	NO		NULL	
SCORE	tinyint(4)	NO		NULL	
idStrategyBasic	varchar(20)	YES		NULL	
idCardSequenceBasic	varchar(20)	YES		NULL	
distToBasic	int(11)	YES		NULL	
idStrategyAdvanced	varchar(20)	YES		NULL	
idCardSequenceAdvanced	varchar(45)	YES		NULL	
distToAdvanced	int(11)	YES		NULL	
changeOfBasicStrategy	tinyint(4)	YES		NULL	
changeOfAdvancedStrategy	tinyint(4)	YES		NULL	
playerStrategySythn	varchar(20)	YES		NULL	

Figure 4.11 A representative database table

The database is a large configuration of tables. There are nearly 5 different databases for the game, each one handling different iterations and different playstyles. There are 11 tables in the database used for the single strategy version of the game. The tables hold information about the users, the games they played, and what strategies they have used. Other databases have a similar set of tables but include tables for aesthetic choices as well. The main table that was used in classifications and for creating the player's fingerprint was the games table. The games table stored information about what cards they had and when they played them, what cards their opponent had and when they played them, how long it took them to play each card in seconds, a time-stamp for when the game was played, what their odds of winning were, what the player's final score was, a number of columns dedicated to discerning whether they developed a strategy or not, and what level of strategy it was, Figure 4.11 shows a breakdown of the table. The reason for using those specific values was that they were needed to know what was played, and in what order, to determine whether a strategy was used and what level of strategy it was.

The ID was needed to determine in what order the games were played. The timestamps help put the games into perspective in case there is a drastic amount of time between games played, and the win percent is a value added for classification purposes to give context to the number of points earned in any given game. The databases were developed using MySQL.

Development tools

Unity 2D

The creation of the game itself was done in the free version of the 2D Unity Engine. Unity was chosen due to its ease of use and large support base. The current version of the game was built in Unity 02017.3.1f1. The game itself is broken up into a number of 2D panels on a canvas that turn themselves on and off depending on the button pressed. Each panel has an on Enable script that runs as soon as it turns on, performing the appropriate task for the panel whether the user knows it or not. The cards used in for gameplay are simple buttons that move and then disable themselves using the animator feature in the Unity engine. There are a number of background scripts that run constantly, such as the connection handler, the options handler and all things related to aesthetic changes, the main canvas, the local storage, and the Auto play for the BOT AI. The game itself had an extensive script that ran several coroutines for timing purposes so that exact timings could be maintained regardless of what the player tried to do during the game. The game has a built-in max time to play a card so that during multiplayer gameplay a player could not be a poor sport towards their opponent by playing slow. This function was maintained in WarCAT to add extra pressure to the player forcing them to think and act faster than they may otherwise want to. The scripts are programmed using C#. Figure 4.12 shows the accreditation for all of the art work used in both WarDOG and WarCAT.

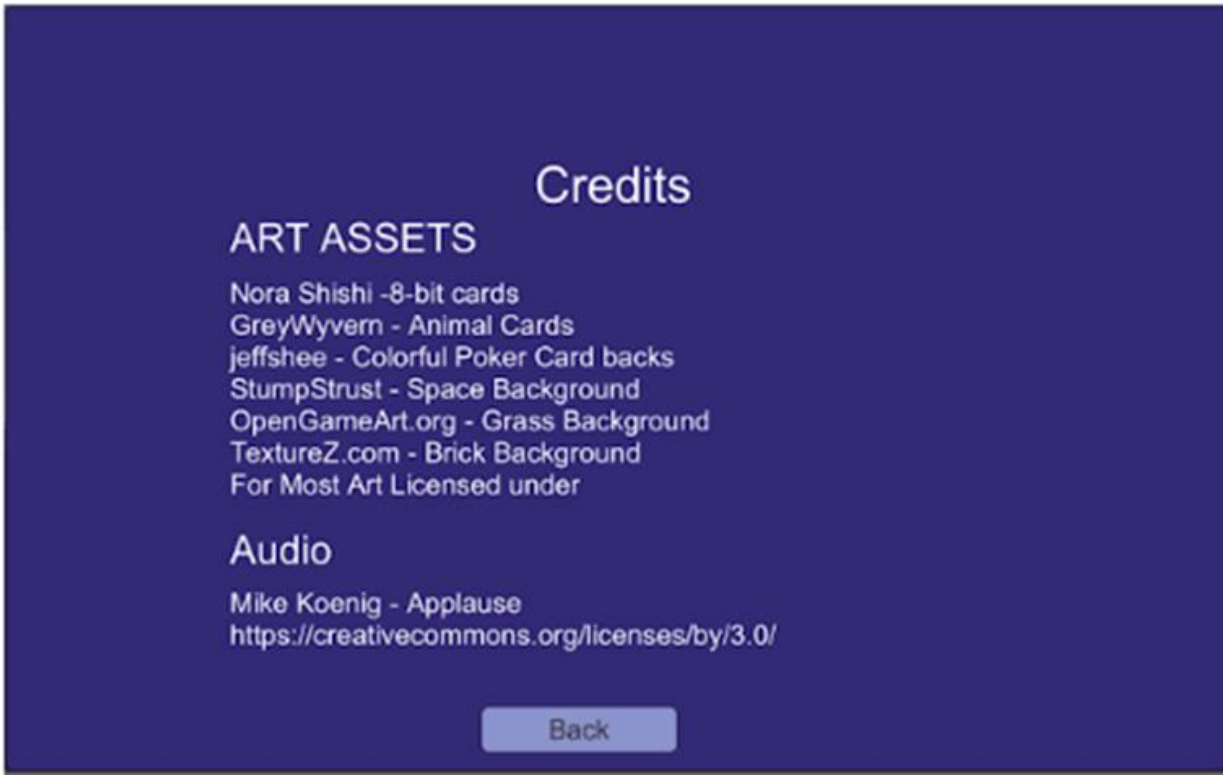


Figure 4.12 the accreditation screen for art assets

Eclipse

The servers were created in Eclipse with the exception of the first version of the WarCAT server and the WarDOG server which were created in NetBeans. The original version ran using NetBeans and did not have a dedicated machine running, instead it was run on the development machine so that the debug and the console could be looked at in real time. The WarCAT server was recreated and improved in the Eclipse IDE with the first version in eclipse being written by my colleague Ben White so that the server was more independent. The server he created was then expanded upon so that it could be more robust and handle all the different versions appropriately. It was also replicated for some of the other versions of the game and was used as reference for the server for the physical sensors.

MySQL Free

The databases were created in MySQL free version. The databases were created using standard practices of the time for what should be keys and how the tables should be linked. The data types for each value in the database were chosen depending on what would take the least amount of space. There is only a single MySQL script created by Dr. Kai Gutenschwager to turn game rows into a strategy value corresponding to the player's level of play. The database section details the breakdown of the different tables.

Methods to potentially detect MCI

The whole purpose of WarCAT is to assess its potential to detect possible Mild Cognitive Impairments. The game has been set up so that we have the potential to add as many testing methods as we need. The simplest method of testing is the game itself. The data that we gather from the time it takes to play each round to the odds of winning all contribute to a fingerprint that can be used to infer whether or not someone has a potential MCI. The more complex methods involve using synthetic data to help increase classification accuracy in the future, at the moment however classification is not done and is being developed by another graduate student in the lab. The various testing methods we have discussed will be discussed in the coming sections.

Base game

The base version of WarCAT for MCI has a number of methods designed to explore whether the game can assist in detecting MCI. The main mode for testing is to gather as much play data as possible and use the idea of emergent strategy. Emergent strategy is the idea that as you play a game, you will naturally develop an optimal play style. In WarCAT, we use this idea to see whether someone is learning how to play the game better and better. Analyzing the data for indications of the player identifying and learning the bot's strategy, and then developing and consistently playing a counter-strategy are hypothesized to correlate to cognitive function or MCI. If someone does have an MCI, we may expect to see a sort of randomness in their play, gaps in strategy, and randomness in their win-loss ratio. To further assess this, we take play data and use it to infer the level of strategy that they are employing from random strategy to advanced strategies. Using these strategy methods, we can form a fingerprint of their playstyle which could then be potentially used to classify cognitive function. The other methods discussed above are to use this fingerprint method in combination with various stressors to exacerbate the

differences in cognitive ability. In addition, it could be possible to employ the gambling versions PvP mode to pit players against each other to see if a new type of emergent strategy would present itself in the form of most effective tactics available or a META. If a META was created, it would allow for even greater differences in data to be formed as in most competitive games that have a META, there are usually tiers of play where players have figured out how best to play the game. In the case of WarCAT, a META would allow us to better classify what an advanced and simple strategy would be so that in the MCI detection play we could more easily create a fingerprint of the players.

However, indications of cognitive lapses in the data cannot be undeniably attributed to an MCI. To determine the actual relationship, a much more controlled and nuanced approach would be needed. The purpose here is to design and instrument a game to assess whether the datasets it generates are of utility in inferring indications of player's executive functions of strategy identification, learning, and implementation against a standardized opponent. Synthesizing new data can assist in this approach. A simple BOT that plays WarCAT against itself was created. The BOT would play an advanced strategy and had a forgetfulness factor where there was a random chance that it would play the cards in its hand from left to right which is effectively random play. The forgetfulness factor was set so that it could be adjusted many times. Other methods for expanding the small dataset include machine learning methods like Reinforcement Learning and Generative adversarial networks.

The next section overviews some of the research outcomes as summarized in several papers.

References

[1] "War – Card Game Rules." n.d. Bicycle Playing Cards. Accessed March 6, 2022. https://bicyclecards.com/how-to-play/war/.

[2] Leduc-McNiven Kyle, Melegrito Genico, Belema Jack-bara, Wurtak Michael. 2018. "Serious Games Development: An Application for Mental Health Assessment." ECE 4600 Capstone Design Project, 2018.

Chapter 5: Academic Contributions

Preface

This section details significant milestones in the work, as shown through publications. My individual contributions to each paper are detailed at the end of the section. The first paper reproduced below was published in the Canadian Medical and Biological Engineering Society Conference (CMBEC) 40, 2017. The paper was co-authored by Marcia R. Friesen, Ian Jeffrey, Kyle Leduc-Niven, Shamir N. Mukhi and Robert D. McLeod. It is included here in its completeness. This was an early formative paper that contained many of the initial ideas as well as our considerable naiveté but helped inform the overall research direction. The second paper was published in CMBEC 42, 2019 and illustrates the progress made in interpreting and visualizing game play data. That paper was co-authored by K. Gutenschwager, K. Leduc-McNiven, M. Aljumaili, R. D. McLeod, and M. R. Friesen. The third paper is from Open Access Text Research and Review Insights, it shows the work done toward the Machine Learning side and was written in between papers one and two chronologically. The paper was co-authored by K. Leduc-McNiven, B. White, H. Zheng, R. D. McLeod, and M. R. Friesen. The final paper was published in International Conference on Artificial Intelligence (ICAI 2019) and discusses the work to develop a cognitive fingerprint based on user strategy before the proper fingerprint display was created. The final paper was co-authored by K. Leduc-McNiven, R. T. Dion, M. Aljumaili, R. D. McLeod, M. R. Friesen, and K. Gutenschwager.

Paper 1: An Innovative Mobile game for detecting Cognitive changes

Marcia R. Friesen, Ian Jeffrey, Kyle Leduc-Niven, Shamir N. Mukhi and Robert D. McLeod
Department of Electrical and Computer Engineering, University of Manitoba

Introduction

This paper presents a mobile game for smartphones and tables, developed to detect the frequency and extent of memory slips. Memory slips are a natural phenomenon associated with cognitive changes as one ages. Although everyone will experience memory slips or more colloquially "senior moments", not everyone who experiences memory slips will develop dementia. However, memory slips however can be symptomatic of more serious cognitive decline and the onset of dementias.

In this work, we present an innovative means to quantitatively assess and monitor memory slips through a non-invasive mobile game.

Background and other methods

A. Background

Currently, over 35 million people are living with dementia and that number is expected to reach 115 million by 2050 fueled by an aging population [1]. The negative impacts of diagnoses on the individual and their family and caregivers are significant, and research implies that early diagnosis can assist in introducing therapies that can slow the progression and maintain quality of life for longer [2].

This work is part of the field exploring the role that mobile devices may play within the niche of mobile apps related to mental health maintenance, assessment and promotion, and specific to the elderly or aging. Mental health apps fall under a much larger space of mobile health apps (mHealth apps) which focus on the practice of medicine and public health mediated by mobile devices. This work falls into the

development of mobile health apps specific to mental health assessment in aging, the process of learning and memory slip, and a technology for empirical assessment. To constrain the scope, this work is contextualized within mobile gaming with the goal of learning a strategy and tracking player performance. Within general health and lifestyle apps, the area of collecting mental health data from Smartphones is very promising [3][4]. Many apps typically address self-monitoring and self-management of mental health concerns such as depression, anxiety, stress, and their associated manifestations. Smartphones offer advantages over other means of mental health technologies [5], such as being very inexpensive, easy to deploy and from which to collect data. While there is little empirical clinical data supporting the efficacy of mental health apps [6], the science is still in its infancy [7] and the landscape points to the need for further evidence-based research [6][8]. Smartphones may have the potential to be effective in many mental health scenarios [6], and while there is research and development activity in this area, few are specifically oriented to aging, Alzheimer's disease and other dementias, or measuring and analysing memory slip [9]. There are other tests for pre-symptomatic Alzheimer's, but currently it is unclear if one test is better—and more likely, several will be used to assess the disease's progression [9]. The research will supplement available tools via long-term game-interaction data collection and its correlation to pre-symptomatic cognitive decline.

B. Other Methods

Alzheimer's Disease does not affect all memory capacities equally: short-term memory (the ability of hold information in mind in an active, readily-available state for a short period of time) is the first to go. Next, episodic memory (memory of autobiographical events) is compromised, followed by semantic memory (memory of the meanings of words and facts about the world). Finally, procedural memory (how to perform tasks and skills) is lost. As the disease advances, parts of memory which were previously intact also become impaired.

Others methods that aim to monitor memory slip and cognitive decline tend to be more deterministic than the techniques described here. For example, for a test at brainhealthregistry.org a person is presented with a series of photos and is asked if they have seen the photo for a second time. This type of instrument addresses short-term memory and does not provide a means for other than binary and deterministic outcomes. Most current screening applications are of this type. On the other end of the test spectrum are eye tracking technologies, premised on the idea that eyes will track in a greater range

and/or at greater speeds when someone is trying to identify an unfamiliar object (i.e. when one believes that one is seeing it for the first time, or is trying to 'place' the image) rather than when one observes a known, familiar object. These are not currently scalable within a Smartphone technology, but also lack any procedural memory component.

Our approach

This project represents one of the first, if not the first, mobile app for purposes of assessment and analysis correlated to memory slips and other age-related cognitive processes, as a means to detect pre-symptomatic cognitive decline. The significance of the research is that, the software framework can be made available to others researchers and to other users, creating a crowdsourced database of game data across multiple age ranges. With a fully functional prototype completed, researchers will be in a very unique position to gain greater insight into player behaviours (cognitive processes) alongside more traditional Smartphone or dementia screening methods.

Previous work within the research group has led to a prototype software platform (mobile game) capable of collecting simple user data. Currently, this framework consists of a prototype online two-player card game, similar to the easy-to-play and familiar card game War (denoted MOBRO WAR). The app deals the player five cards. All five cards are known to the player ('face up') but are not known to the app (a 'bot' acting as the opposing player). The player plays a card and the 'bot' plays a card in return (Fig. 1); the higher of the two cards wins the hand. A single round consists of five hands and is completed in approximately 15 seconds. Each player knows their own cards and can play them in any order during the round. In its current form, the 'bot' plays the same strategy for 100 rounds, either low-to-high, high-to-low, or middle-high-low, and it is incumbent upon the player to recognize the 'bot's' strategy and counter appropriately. At each round, the probability of winning by chance vs. winning by learned strategy are recalculated, and the player's score is evaluated accordingly.

Further work will focus on developing this prototype into a robust and reliable software platform capable of collecting data related to player performance useful to assessing strategy learning and retention. The collected data will include duration of play, speed of each round, number of games played, and player choices in each card hand. The data will be collected to determine differences between older and younger players as well as between various other groups.

Figure 1: Example play of MOBRO WAR

There are several design factors that support the conjecture that this instrument can contribute to quantitatively assessing memory loss:

- Our mobile app uses engages short term, episodic, semantic as well as procedural memory. That is, our app requires both recall as well as thinking/analyzing in determining or learning a strategy as well as recalling strategy.
- Since the game is played repeatedly in short burst, a person can learn a strategy and, in the absence of cognitive decline or memory slip, will tend to reinforce their strategy with an increasing score differential.
- The game involves a degree of randomness in that the cards are dealt from a deck randomly for both the player and the 'bot', and each hand's five cards are unlikely to be the same as the five cards of the previous hand. Because of this random element, it is possible to play the learned strategy, i.e. the winning strategy (on average) but still lose a round or hand, and this can bring on a moment of confusion and trigger a memory slip. Since the player's strategy will be recorded, one will be able to differentiate a strategy used consistently periods in which a person may have forgotten their strategy or is not using one. Both outcomes impact the calculated probability of winning by more than chance; however, since one will know the cards played, we will be able to infer whether or not the person developed and maintained their strategy. Further, if the memory slip persists, this again may be measurable through a delay in returning to a winning strategy.

- Each round is played with random cards which also can induce a moment of rethinking that can bring about a memory slip. These latter points are somewhat speculative but worth further research.

In effect, the randomness of the cards and the probabilistic event of losing the hand even though the "winning: strategy was played both have an impact. These features make this mobile game different from others which tend to be more deterministic.

To accommodate players who would find a five-card hand too difficult (for example, due to a higher degree of cognitive decline), the game can easily be redeveloped with only three cards, thereby reducing the apparent degree of difficulty while retaining the random elements described above, which are essential to the game's novelty and originality in potentially triggering and monitoring a memory slip.

At present, there is very little research on real-time data collection within mental health games. This work represents the first effort to capture player behavioral patterns with emphasis on better understanding memory slips, widely recognized in the literature as a potentially major step in assisting in assessing cognitive degeneration. What follows are some theorized examples of how the collected data may be used to better understand the analytic potential.

Understanding demography: Collected data could be used to explore the demographic profile of participants in the game/s. We anticipate that analysis by age, sex, region (potentially taken by proxy measures of IP addresses) could be performed. However, there are ethical and data confidentiality issues to be addressed, and it is uncertain whether this could be achieved in practice and beyond the scope of the present paper.

Understanding different types and patterns of play: We anticipate that sub-group analysis will be possible. Having complete control over the platform, it will be possible to gain objective data about frequency of play, length of sessions and player choices in each card hand, which would be informative in describing and contrasting different types of participants.

Understanding trajectories: If an on-going relationship can be established with a group of participants, a system can be set up to explore the behavior and the play of select individuals longitudinally, examining changes in their play strategies and scores over time. It could also accommodate examination of behaviours of groups over time, including those who play consistently at certain levels, those who increase their involvement with these games, those who display variable patterns of play and those who stop play. Once these trajectories are identified, their profile may be explored.

Most importantly, the player's user data allows us to assess their probability of maintaining a winning strategy throughout play. As a person encounters difficulty with recall, the probability of winning is reduced until the participant recalls or relearns their winning strategy.

Table 1 illustrates some preliminary results from a simple user trial illustrating some of the data currently collected by the mobile game, illustrating a person's ability to beat a bot by more than chance (statistically significant).

Table 1: Player records (prototype test volunteers

User	Level 1		Level 2		Level 3		Level 4	
	win	loss	win	loss	win	loss	Win	loss
Participant 1	142	53	220	74	292	101		
Participant 2	70	30						
Participant 3	85	12	160	34	219	50		
Participant 4	80	16					317	72
Participant 5	69	28	208	87				
Participant 6	80	19	162	35	256	51		
Participant 7	75	26						
Participant 8	76	23	149	50	213	83		

For development iterations, we control installations of the mobile game on to users' smartphones by hosting the game app ourselves on a secure server and inviting volunteers and the research team to "side load" the game. This allows considerable control over the game during development and debugging.

One would be remiss not to mention the opportunities for integrating machine learning (ML) into mental health Smartphone apps. Such applications include ML approaches to assessing speech and language disorders from speech samples [10] and mental health assessment through face image classification [11].

Summary

This paper presented a novel means of assessing memory slips, associating them with cognitive decline. The statistical method tracks a person's ability to learn and retain a winning strategy in a mobile game over time. If a memory slip is detected or a player's

pattern of play varies significantly, it may be as a result of a memory slip in comparison to their more normal play. The mobile game (app) will be demonstrated at the conference, including the opportunity for others to play several rounds. Perhaps somewhat serendipitously, one may experience a memory slip while playing.

References

[1] World Health Organization and Alzheimer's Disease International, *Dementia: A Public Health Priority,* WHO, Geneva, Switzerland, 2012.

[2] M. Prince, R. Bryce, and C. Ferri, *World Alzheimer Report 2011: The benefits of early diagnosis and intervention*. Alzheimer's Disease International, 2011. Available: https://www.alz.co.uk/research/world-report-2011.

[3] D.D. Luxton, R.A. McCann, N.E. Bush, M. Mishkind, and G.M. Reger, "mHealth for mental health: Integrating Smartphone technology in behavioral healthcare," *Professional Psychology: Research and Practice*, vol. 42(6), pp. 505, 2011.

[4] J. Torous, R. Friedman, and M. Keshavan, "Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions," *JMIR mHealth and uHealth*, vol. 2(1), e2, 2014.

[5] J. Proudfoot, "The future is in our hands: the role of mobile phones in the prevention and management of mental disorders," *Australian and New Zealand Journal of Psychiatry*, vol. 47(2), pp. 111-113, 2013.

[6] T. Donker, K. Petrie, J. Proudfoot, J. Clarke, M.R. Birch, and H. Christensen, "Smartphones for smarter delivery of mental health programs: a systematic review," *Journal of Medical Internet research*, vol. 15(11), e247, 2013.

[7] E. Anthes, "Mental health: There's an app for that," *Nature News*, vol. 532, 2016.

[8] C. Hollis, R. Morriss, J. Martin, S. Amani, R. Cotton, M. Denis, and S. Lewis, "Technological innovations in mental healthcare: harnessing the digital revolution," *The British Journal of Psychiatry*, vol. 206(4), pp. 263-265, 2015.

[9] E. Landhuis, "Catching Alzheimer's before Memory Slips," 2015. Available: https://www.scientificamerican.com/article/catching-alzheimer-s-before-memory-slips/

[10] I. Song and J. Diederich, "Speech analysis for mental health assessment using support vector machines," In *Mental Health Informatics* (pp. 79-105). Springer, Berlin, 2014.

[11] R. Wang, A.T. Campbell, and X Zhou, "Using opportunistic face logging from Smartphone to infer mental health: challenges and future directions," *Proceedings of*

the 2015 ACM International Symposium on Wearable Computers, pp. 683-692), 2015.

Paper 2: Visualizing "Cognitive Fingerprints" from Simple Mobile Game Play

K. Gutenschwager[1], K. Leduc-McNiven[2], M. Aljumaili[2], R.D. McLeod[2], and M.R. Friesen[2]

[1] Institute of Information Engineering, Ostfalia University of Applied Sciences, Wolfenbüttel, Germany
[2]Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Canada

Abstract— **Serious Games and associated data analytics offer the potential of a complementary means of detecting early signs of mild cognitive impairment (MCI), which is often a precursor to more serious forms of dementias. As with all diseases and illnesses, the ability to mitigate the impact of the illness is directly correlated to early detection and intervention. In this work, a representative serious game is used to capture a "cognitive fingerprint" of a person's play, which is then used to analyze and visualize play. The long-term objective of the research is to demonstrate that data collected from serious games may be used to detect cognitive difficulties that may be pre-symptomatic, and outside the scope of normal age-related cognitive decline. The present work assesses the viability of the platform for this purpose and opportunities in data visualization, but does not include clinical testing for MCI.**

Keywords—***Serious games, mild cognitive impairment, machine learning, data visualization.***

I. INTRODUCTION

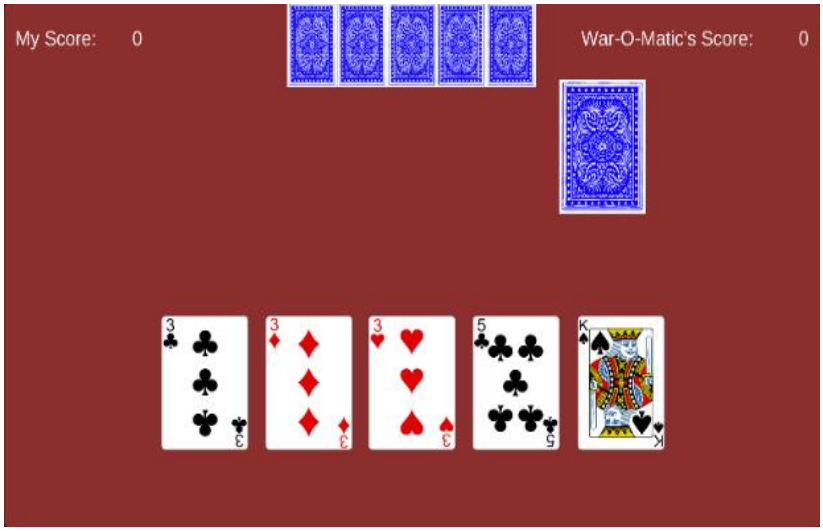
It is increasingly apparent that serious games are a becoming rich source of data associated with many aspects of mental health. Other researchers have surveyed current research and development of serious games for mental health data [1][2]. While there are other games on the market that claim to do 'brain training' to maintain cognitive function or potentially delay MCI, no game has been developed to *detect* MCI it – a subtle but important distinction. This work focuses on analyzing data from games toward MCI detection [3].

This paper briefly describes a serious game developed within the research group, denoted WarCAT, which illustrates a game's ability to capture a player's cognitive function related to strategy development, memory, recall, and other executive functions.

We then briefly present an initial attempt to classify play using synthetic agents playing with various degrees of impairment. Limited success using various machine learning techniques prompted a more direct statistical approach of inferring cognitive ability by analyzing specific strategies a person may use or develop based on detailed moves made within the game. This process provided a data visualization of a person's "cognitive fingerprint" of play. The data visualization and rules of inference associated with play will be presented in detail with examples extracted from a small tournament.

II. WarCAT

WarCAT is a simple mobile game based on the familiar card game of WAR. Our variation is that a hand of 5 cards is dealt to the player, who can play them in any order they choose. The person is playing against a "bot" or machine that has also been dealt 5 cards (unseen by the player). The player and the bot both lay a card, and the higher card wins. The bot's strategy of play is to play its cards one at a time in descending order. There are other variations of bot strategy, but for the purposes here, this particular strategy is sufficiently representative. The player's challenge is to recognize that the bot is playing a strategy, and to consistently counter the strategy in one's own play.



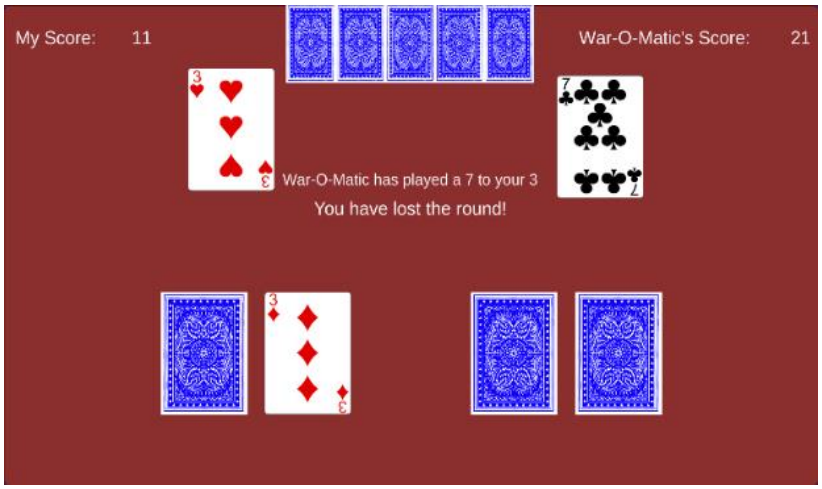


Fig. 1 Screenshot of one hand and instance of play.

Figure 1 illustrates a hand that was dealt and an instance of play. The player selected a 3 as the fourth card played against the bot's fourth card which was a 7. In this case, the bot won the hand. A game lasts for 50 (settable parameter) hands of play, at which time a player levels up. There are a programmable number of levels but for purposes of discussion, three will suffice. In the base version of the game, the bot does not change its strategy at subsequent levels. Other variations of the game have both changing strategy at subsequent levels, e.g. playing its cards in ascending order or other variations. At this time, as much player data as possible is collected, including the card played for each hand on both sides, whether the player won or lost the hand, the score, and timing information. From this, considerable information can be inferred as to whether (and how quickly) the player recognized and countered the bot's strategy, remembered the strategy to consistently beat the bot (when possible), or experienced points of confusion or momentary "loss of set" – i.e. forgetting the strategy, and then re-learning it (or not) at some later point in the game. As the cards are dealt stochastically, there is an inherent potential for confusion, for instance as a consequence of knowing that one played a winning strategy yet still lost the hand – which is a situation that happens approximately 25% of the time with the scoring system currently in use and playing a "good" strategy to beat the bot.

The scoring system scores a narrow margin of victory more significantly than by beating the bot by a wide margin. The score for each card played with a hand is:

$$score = 13 - (Winning\ card\ value - Losing\ card\ value)$$

unless there is a tie, in which case the score for that play is 0. This also supports strategy learning through reinforcement.

By contrast, a simple scoring system would award a point to the player or bot depending upon who simply had the higher card. The difficulty with simple scoring is that it is more difficult to play a winning strategy and more consistently win. With simple scoring, a person can expect to lose (as a consequence of the stochastic nature of the cards) approximately 35% of the time. It is easier to learn a winning strategy through reinforcement if one wins more often.

III. Machine Learning Forays

Initial attempts to use Machine Learning (ML) to classify play used synthetic input data generated by bot vs. bot play (computer agents), where one bot played at various degrees of impairment to mimic a human player with none-to-some MCI, while the other bot played their strategy consistently. Data were generated by bots that would play either a winning strategy or a random strategy governed by Bernoulli trials. For example, a bot would play at random with a probability p and a winning strategy with a probability of (1-p). The impairment model was improved to include impairment of a bursty nature using a process for modeling bursty channels [4].

The ML techniques investigated using synthetic data included dense neural networks (DNN) and convolutional neural networks (CNN) within the TensorFlow framework [5]. In both cases, the classification accuracy after training was in the mid-90%. Effectively, these results are a result of an accurate inference of whether the play was impaired or not. Having an accurate inference of whether the play was impaired of not would lead to similar (a few % less accurate) classification from simple statistics alone.

IV. Inferencing and Data Visualization

These observations lead to a more deliberate approach on attempting to differentiate impaired play from a player deliberately choosing an adaptive strategy with respect to the cards played by the bot. Even though WarCAT is simple and apparently straightforward, interpreting play (confused vs. adaptive) is difficult. This difficulty would be compounded if the game were any more complex.

A. Re-Interpreting Strategy

Much of the following is based on actual play and observations made during play, which has led to re-examining a person's optimal strategy. Knowing that the bot is always playing high to low cards, a reasonably good strategy which one could already consider as "perfect play" is to consistently play one's lowest card first followed by highest to lowest. In this strategy, the first card is sacrificed to see the bot's highest card. This strategy leads to wins approximately 75% of the time, but doesn't account for more subtle decision

making during the hand. However, there are other strategies which may also lead to equal or better results. In order to evaluate the 120 possible basic strategies, defined as the possible permutations of the five cards, we performed a simulation experiment where 10M games were played against a bot that always plays its cards high to low. We further analyzed the results by storing information about the lowest and the highest card value, as more advanced players might use this information as input to a more adaptive strategy. Our results show that looking at all 78 possible combinations of lowest and highest card, 23 different permutations of cards lead to the best results at an average for at least one combination of lowest and highest card.

TABLE I: Winning strategies with respect to the lowest and highest card. Note that other strategies are minor variations.

ID	Sequence	Description
1	**1**-**2**-5-4-3	Sacrifice first two lowest cards at first and second position.
2	**1**-5-4-3-2	Sacrifice lowest card at the first position.
3	**1**-5-**2**-4-3	Sacrifice first two lowest cards, but the second best card at third position.
4	5-4-3-2-1	Play high to low
5	5-**1**-4-3-2	Sacrifice lowest card at the second pos.
6	**2**-**1**-5-4-3	Sacrifice first two cards (second best first)
7	5-4-3-**1**-2	Sacrifice lowest card at the forth pos.
8	5-4-**1**-3-2	Sacrifice the lowest card at the third pos.

If a player always plays an adaptive strategy, then the respective sequence to play their cards based on the present combination of the lowest and highest card (in their hand) leads to the percentage of wins increasing to about 83.4%. The main eight of the 23 strategies are given in Table I.

If we reduce the number of different strategies to four (IDs 1, 2, 4, 5), we obtain a rather manageable set of strategies and combinations of lowest and highest cards and when to use them. The percentage of games won is still very high with a value of about 82.7%.

Note that if a player also takes into account the first card played by the bot and then adapts their strategy based on that card, the chance of winning further increases by about 1%.

Looking at the limited number of games a player completes in one round, the difference between these two approaches is hardly recognizable and statistically insignificant. We therefore differentiate a few main strategies that actually lead to statistically significantly different results which a player may actually recognize playing only a few rounds.

Table II estimates the expected points awarded based on the strategy played. For example, if a player played at random, the average number of points expected would be 0. If they developed a strategy of consistently sacrificing their first card followed by playing their cards from high to low, their expected points per hand would be 16.6. Within one level, a person would play 50 hands, and if they proceeded to play that strategy, they would end up with approximately 50*16.1 points.

TABLE II: Expected points per hand, based on strategy played

ID	Strategy	Avg. points
1	Random	0
2	High to low	0
3	Low to high	0
4	Sacrifice first card, rest high to low	16.6
5	Sacrifice first two lowest cards, rest high to low (lowest at pos. 1)	16.1
6	Adaptive (1): If the value of second lowest card < 8 then sacrifice two cards, else one	18.1

Overall, we can vary the strategy a player follows as the first step, and then analyze if the player gets distracted from that strategy over time. However, there are a number of complexities that need to be taken into account. A player may tend to deliberately change

strategy over time to improve their overall score. These variations make it more difficult to infer moments of confusion or loss of set.

The first three strategies can be differentiated by the sequence played, even though the average score does not differ. One can use a simple distance measure for comparing two sequences played, denoted as P1 and P2, where pos_j states the position of the j-lowest card in the sequence:

$$dist = \sum_{j=1}^{5} abs(pos_j^{P1} - pos_j^{P2}) \quad (1)$$

Computing the average distance for the last n games to the (best) sequence of the respective main strategy (2-7), one can analyze the current distraction as a deviation from the optimal sequence of the strategy the player follows momentarily (except for random play). Taking into account the average score obtained using a specific strategy, one can further compare the average score the player achieved with the respective expected value. This information might give further insights into why a player changes strategy, i.e., if the actual score is lower than the expected value (computed by our simulation experiments), there might be a higher probability that a player got distracted from the chosen strategy.

The classification on whether and how significantly a player's behavior reflects a cognitive impairment needs to be corrected by this influence. Furthermore, one needs to take into account a learning curve, such that we compare a player's current play to the best strategy learned so far. This means that the "perfect play", as introduced above, is actually to be considered as dynamic if one wants to judge the influence of cognitive decline.

B. Visualization of a Cognitive Fingerprint

In order to better understand the patterns of play, a technique was developed to visualize an inferred strategy, duration of play at that level, and overall score achieved during play (see Fig. 4 for further explanation).

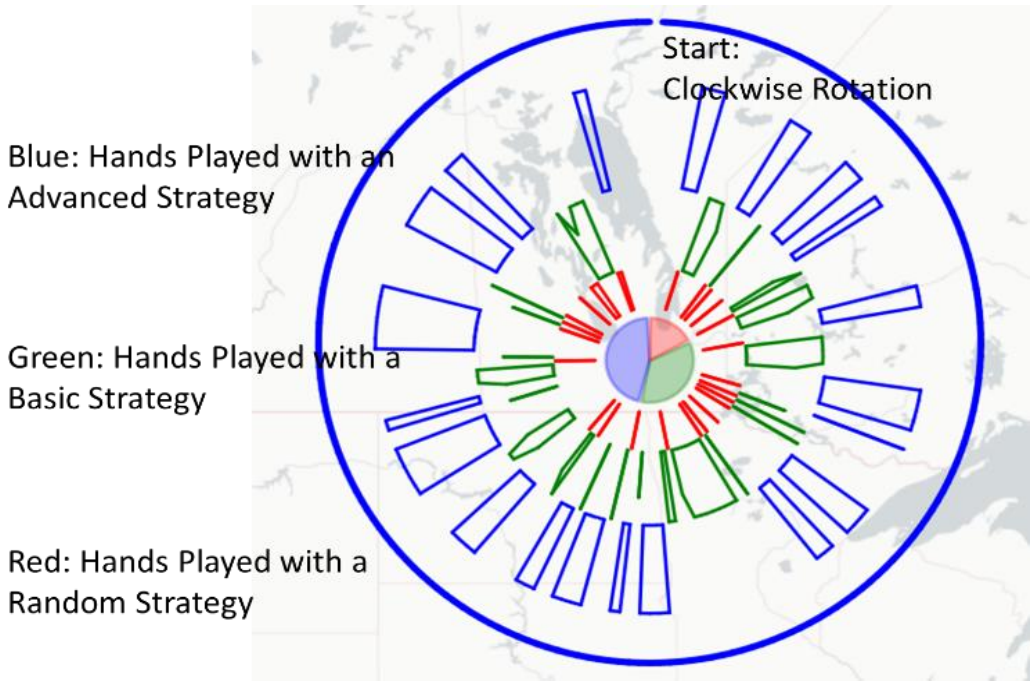


Fig. 4: Visualization of winning play

To illustrate this with real data, a small tournament was organized early December 2018 to play for fun with an incentive of the high score receiving a $100 prize. This was an informal tournament that allowed the collection of a small player pool of anonymous data. The objective was simply to explore the visualization of data extracted from play. The tournament consisted of games of 3 levels of play where the bot would play its cards from high to low at each level. Each level consisted of 50 hands. Players were allowed to play multiple times. University research ethics approval was obtained to assess the game hedonics, software verification, and to illustrate the interpretation and visualization of gameplay. The tournament was not carried out to assess MCI.

Based on the inferences of Section IV(*A*) the data was displayed on a polar plot with coordinates governed by the time of the hand and an inference of the strategy of play. Figure 4 illustrates play by the winner of the tournament. In this case, most of the play was inferred to be Advanced, with only a small portion inferred as Random or Basic.

Another example of play is visualized in Fig. 5. The player was an adolescent who had never seen the game previously. Various strategies of play can be seen as the person quickly developed more advanced strategies.

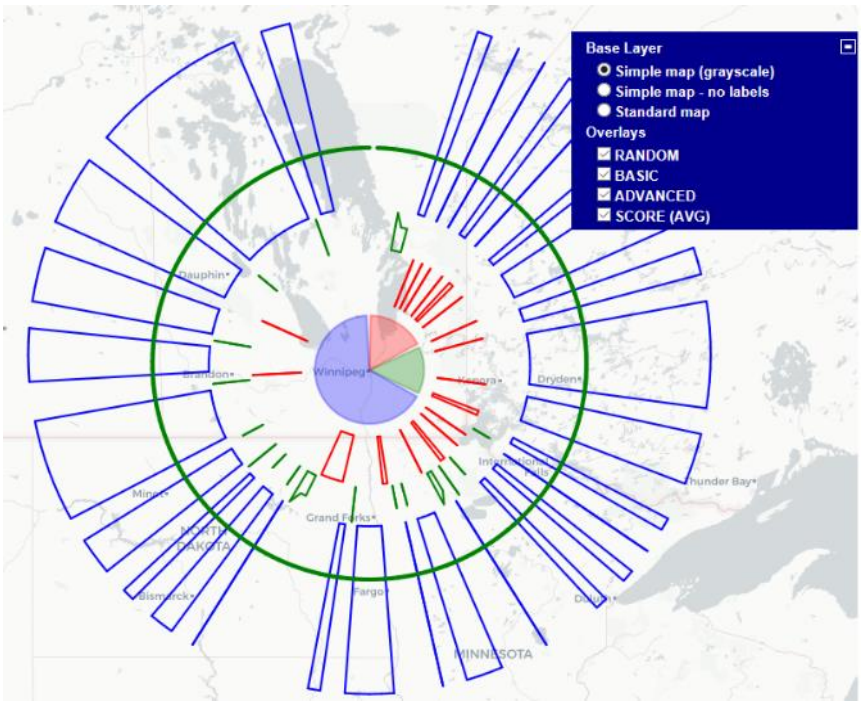


Fig. 5: Visualization of novice play

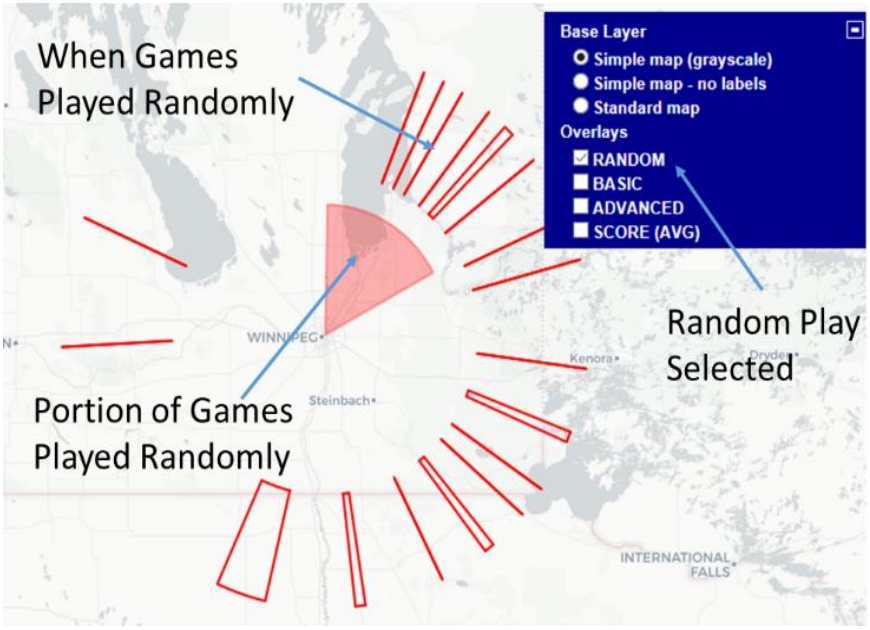


Fig. 6: Visualization of random play during game

Figure 6 illustrates a selection of inferences made of Random or Basic strategies of play, and these features that are conjectured to be useful as ML inputs. As with many ML applications, data preprocessing is essential in obtaining useful and accurate predictions. Figure 7 illustrates a visualization of the best play seen to date.

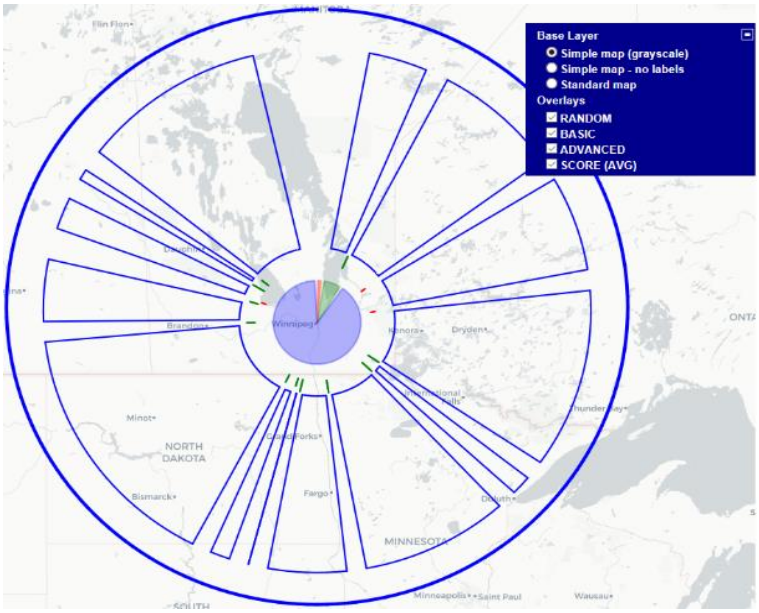


Fig. 7 Visualization of best play to date

V. CONCLUSIONS

Building serious games for MCI detection and attempting to extract players' patterns of play to assess cognitive approaches during play is very difficult. Even at the level of our very simple game, analyzing data is complex. These complexities will be compounded by the subtleties of MCI and the real (future) challenges of labeled data. Future work will be in developing agents or bots that will learn to beat the game consistently, for the purposes of emulating human play. Depending upon the degree of agent training, it may be possible to label various levels of play or impairment. These types of synthetic data will be visualized with the methods developed here as well as with CNNs for more accurate predictions of cognitive impairment distinct from normal cognitive decline.

VI. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

VII. REFERENCES

1. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform. 2012 Jan;177(2012):85-96.
2. Muscio C, Tiraboschi P, Guerra UP, Defanti CA, Frisoni GB. Clinical trial design of serious gaming in mild cognitive impairment. Frontiers in aging neuroscience. 2015 Mar 11;7:26.

3. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Archives of neurology. 1999 Mar 1;56(3):303-8..
4. Elliott EO. Estimates of error rates for codes on burst-noise channels. The Bell System Technical Journal. 1963 Sep;42(5):1977-97.
5. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

Paper 3:

Serious games to assess mild cognitive impairment: 'The game is the assessment'

Kyle Leduc-McNiven, Benjamin White, Huizhe Zheng, Robert D McLeod* and Marcia R Friesen

Department of Electrical & Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

Abstract

Background: Early recognition of mild cognitive impairment (MCI) and subtle changes to cognitive abilities that precede an MCI diagnosis has the potential to improve the efficacy of therapeutic treatment programs.

Objective: The work addresses mobile games' potential as empirical assessment tools for cognitive processes within the domains of attention, recognition, recall, and memory applied to game strategy.

Methods: Two games have been developed with this objective. WarCAT is based on a familiar card game, War, and "Lock Picking" is a search for an optimal score, akin to finding the combination that opens a lock. Both games provide players with immediate feedback but engage different algorithms and heuristics to solve the respective problems at hand.

Conclusions: By collecting player data on large scales to allow for baseline establishment of cognitive abilities across demographic (age) profiles, longitudinal performance of individuals and of groups can be established, and from there, the potential exists to employ machine learning methods to detect subtle changes in an individual's cognitive processes over time.

Received: February 17, 2018; **Accepted:** March 19, 2018; **Published:** March 23, 2018

Abbreviations: ANN: Artificial Neural Network; BCI: Brain-Computer Interfacing;

CNN: Convolutional Neural Network; ML: Machine Learning; MCI: Mild Cognitive Impairment; ROC: Receiver Operating Characteristic; RL: Reinforcement Learning; WarCAT: War Cognitive Assessment Tool.

Introduction

Mild cognitive impairment (MCI) constitutes a clinical entity differentiated from healthy control subjects and those with very mild Alzheimer's Disease [1]. Early recognition of MCI and subtle changes to specific cognitive abilities that precede an MCI diagnosis has the potential to improve the efficacy of therapeutic treatment programs [2,3]. Currently, over 35M people worldwide live with dementia and that is expected to reach 115M by 2050, fueled by an aging population, as aging is the biggest risk factor for cognitive decline and dementia [4]. The negative impacts on the individual, family, and caregivers are significant. As disease-modifying treatments are discovered, early diagnosis will be essential to assist in introducing therapies that can slow the progression and maintain longer quality of life [5].

Mental health apps are part of a much larger mobile health (mHealth) space, and the work addresses mHealth apps' potential as empirical measurement tools for cognitive assessment including memory, learning, problem-solving, and other executive processes. The work finds clinical relevance in the early identification of cognitive decline, crucial for optimal pharmacological treatment and timely provision of psychosocial care [3,5-9]. The first symptoms of cognitive decline may be present several years before a clinical diagnosis of dementia can be made and thus tools that detect underlying patterns of brain dysfunction are of great importance.

This work combines mHealth with MCI assessment. This work goes beyond developing electronic/mobile presentations or reproductions

of existing MCI assessment instruments; rather, the work presents mobile games that inherently provide a cognitive assessment function via the analysis of an individual's game-playing data.

Clinicians indicate that standard MCI assessment instruments (non-electronic) can, at times, induce anxiety in the individual who knows they are being tested and may become frustrated at their self-perception of performance on the assessment. While there is a significant market of gamification of cognitive stimulation ("brain games" like Lumosity), there is currently little gamification of MCI assessment as one means to address this anxiety [10,11].

Gamification of MCI assessment also open up opportunities to collect player metadata on large scales that allow for baseline establishment of cognitive abilities across

demographic (age) profiles, longitudinal performance of individuals and of groups, and from there, the potential to detect subtle changes in an individual's cognitive processes over time. It is the self-perception of losses of specific cognitive processes such as recognition and recall that can cause anxiety to individuals being assessed. The proposed tools have been designed to include the ability to objectively assess recognition of a game strategy, recall of the strategy, failure to maintain set (reverting to a different strategy or no strategy at all), and perseveration (reverting to an earlier strategy).

The games outlined below are exemplars of a genre of tools that have the inherent potential to provide data on cognitive functions. One is based on a familiar card game, War. In this variant, a person plays iteratively against a bot (computer) that is playing a consistent strategy. The person is provided immediate feedback of each round of play, allowing them to learn a winning strategy through reinforcement learning. Play continues against the bot until the person has demonstrated that they have won by more than chance, at which point the player "levels up" to continue. In the next level, the bot consistently plays a new strategy. By the stochastic nature of the game itself, there is opportunity for distraction, facilitating temporary lapses of concentration or memory. As all moves are tracked, there is an opportunity to analyze play in considerable detail via the metadata, essentially tracking "memory slips" in real time. The second game is an analogy to opening a barrel-type combination lock. In this game, rather than recognizing and countering a consistent strategy, the player is challenged to come up with a strategy to open a combination lock with play-by-play feedback of how close they are to "picking" the lock. Both games provide real-time data related to a person's play.

WarCAT for MCI assessment

WarCAT framework

WarCAT (War Cognitive Assessment Tool) is one of the games developed. The following is a description of the game and how it can be used as a cognitive assessment tool with qualitative as well as quantitative outputs.

The underlying algorithmic problem is more complicated than a search and is more closely related to the problem of sorting. Sorting is part of many computer as well as human problem-solving methods. Overall, people do need to learn sorting generally, and most acquire that ability early on in their cognitive development. People are reasonably good at sorting, although more likely akin to selecting an item from the unsorted group followed by insertion into a sorted group. For example, with respect to playing cards, many people have varied techniques or implicit algorithms that work well with small

numbers of cards. The reason most approaches work well is the small problem instance size (N, the number of cards), and if one were to analyze the complexity of the algorithms a person used to sort them, there would be little difference as we are nowhere near an asymptotic limit.

The framework for WarCAT takes the form of a mobile game, an unexplored genre for MCI assessment. In addition to being mobile, it is also concurrent (head-to-head) and competitive. In its present state, each game consists of five rounds of the card game War, each game played is measured in seconds, and feedback is near instantaneous. Presently, three levels of a minimum of 100 games each have been implemented in the prototype, and the human player plays against a bot. The bot maintains a constant strategy for at least 100 games, and only after a player has demonstrated that they have beaten the bot by a non-chance margin can the player 'level up'.

WarCAT combines real time player behavioural measures captured via metadata (e.g. play by play moves, duration played, duration per move, number of moves needed to level up, etc.) as well as self-report measures. The basic gaming framework has been further developed during the summer 2017 to optionally include more traditional cognitive assessments such as a paired associated learning test as an in-app feature which the player completes to proceed.

Figure 1 shows a screenshot of the proof of concept, which is currently rudimentary but functional. The registration screen illustrates the demographic data requested (optionally).

During play, the user is provided immediate feedback as to having won or lost the game (Figure 2). Figure 3 illustrates the feedback provided to a player when the player has beaten the bot by more than chance (statistically).

Also illustrated (Figures 3) is the aggregate strategy that the player played. For this instance, the person played mid-level cards followed by low cards followed by their high cards. This corresponds to the strategy they recognized early in the level, in which the bot consistently played low-high-medium cards. As the game collects extensive hand-by-hand data, one can infer a person's ability to learn their winning strategy as well as periods where they may have forgotten their strategy or had difficulty in arranging their cards for play. This type of ordering cards presents challenges, even more so as one's cards are selected at random.

WarCAT: Opportunities in data analysis

Visualizing a trajectory of learning a winning strategy would be very difficult.

However, by virtue of the fact that the game is stochastic, there is opportunity to analyze a player's behavior through the use of the corresponding confusion matrix and Receiver Operating Characteristic (ROC) curves for classification. Fig. 4 illustrates data collected per hand. It illustrates a phase of learning and forgetting and recall or relearning. Figure 4 provides a fingerprint of a person's play suitable as input into a classifier.

The research challenge is in using player metadata for analysis in confusion matrices, ROC curves, and clustering and/or classification of play through machine learning (ML) techniques. Once sufficient data are collected, an artificial neural network (ANN) can be trained with the data to classify the degree to which a person experiences cognitive difficulty during play. Another benefit is being able to monitor play behavior longitudinally. The ML classifier would be set up as follows. Initially bots will play against bots. During play, by virtue of the stochastic nature of the cards being dealt, a given hand will either be a true positive (Win when one should win as well as lose when one should lose) in contrast to the case where one would lose even though they played their winning strategy (i.e. dealt bad cards).

Specifically, Figures 4 and 5 illustrate a fingerprint of a person's play (in this case, RDM). The differences in color for brief periods illustrated where RDM's strategy varied from the optimal. The evidence is somewhat anecdotal as RDM would experience long periods of having learned and implemented the strategy, followed by periods of "memory slip". These fingerprints would be used in a classifier to attempt to group differences in play.

A challenge to ML analysis is that currently, data are only labeled with a person's age and gender which they agree to provide when registering to play (Figure 1), designed explicitly as minimal and noninvasive data to collect. The objective is to have a sufficient number of people of all ages and of healthy cognitive function play to be able to correlate play behavior to age and potentially to gender.

A perfect game by a "player" bot would be consistently playing the winning strategy against the "house" bot. For example, if the house bot were consistently playing low to high, the player bot would sort their cards and play them in an offset to the house bot. If the house bot played their lowest card first, the player bot would play their second lowest card, etc. With this scenario the probability of winning a 5 card hand is approximately 66%. This is confirmed semi-analytically as follows.

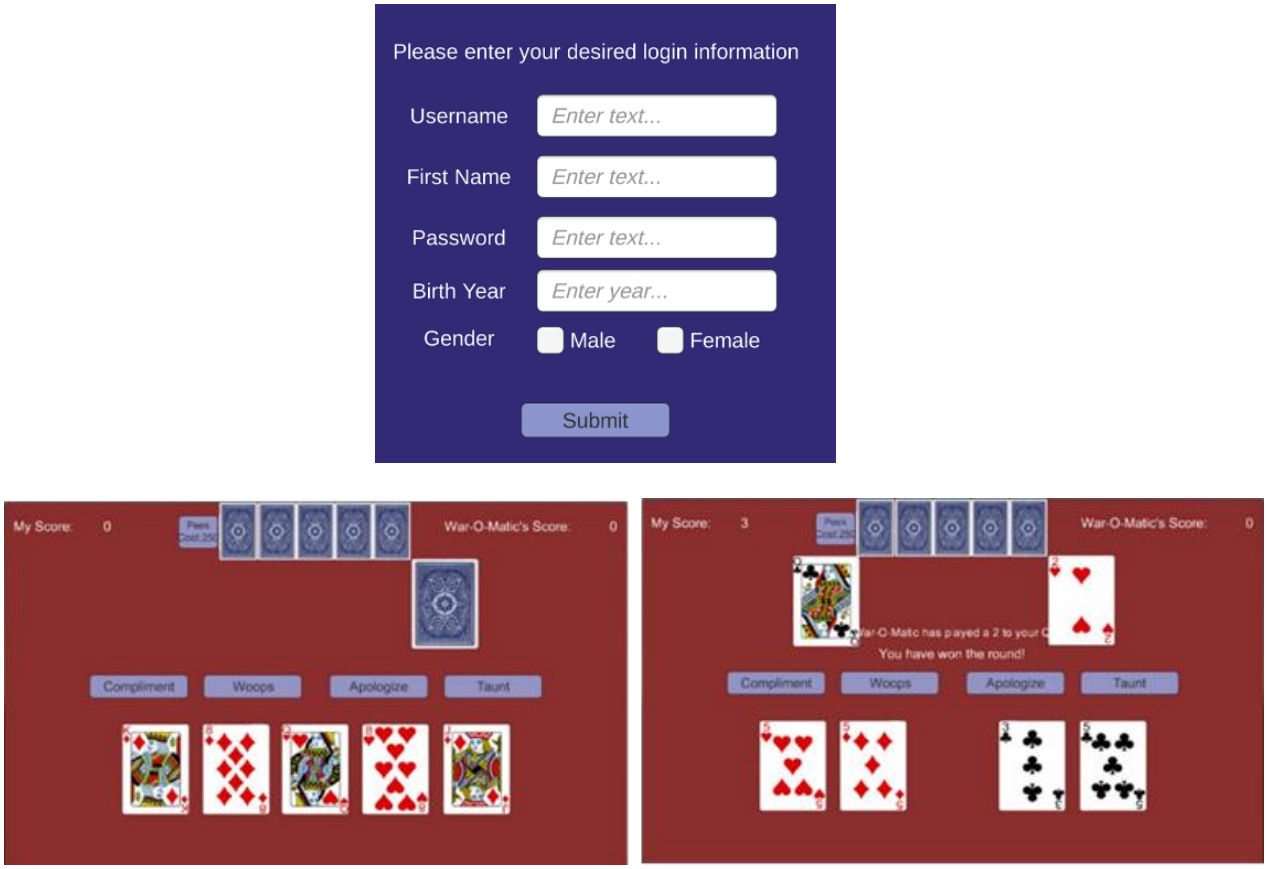


Figure 1. Example play of WarCAT

Figure 2. An example of feedback per round (reinforcement learning)

Figure 3. An example of successfully beating the bot through three strategies. More Fun: Fireworks and a Dancing Cat (Dancing Cat not yet implemented)

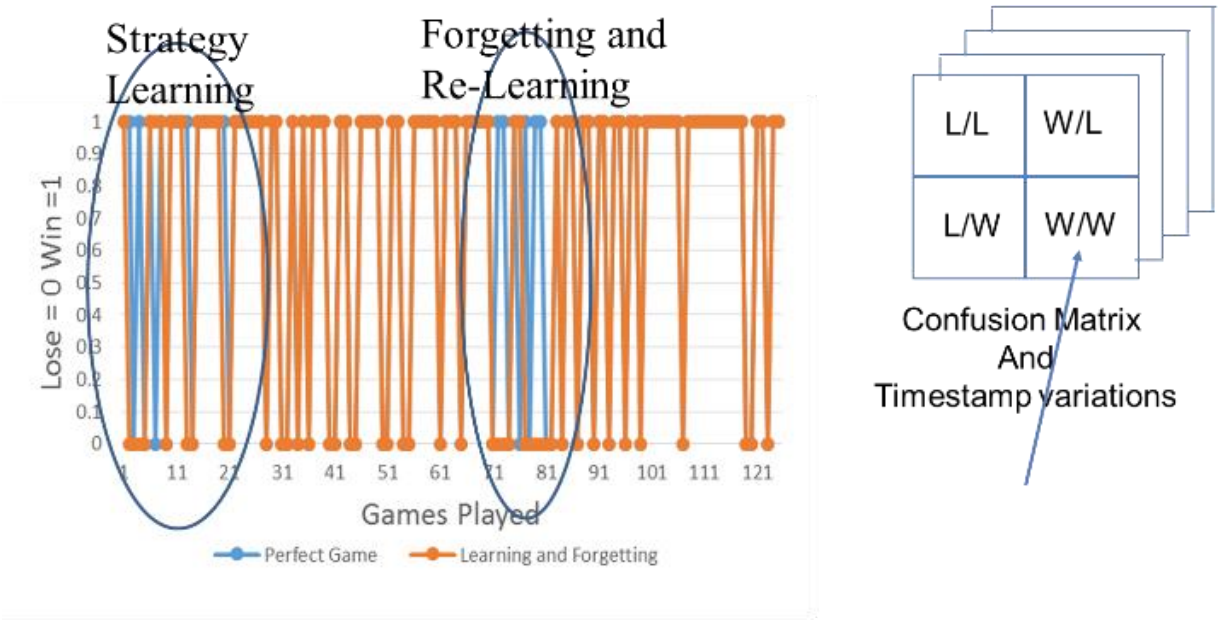


Figure 4. Illustration of data collected through a whole level of play

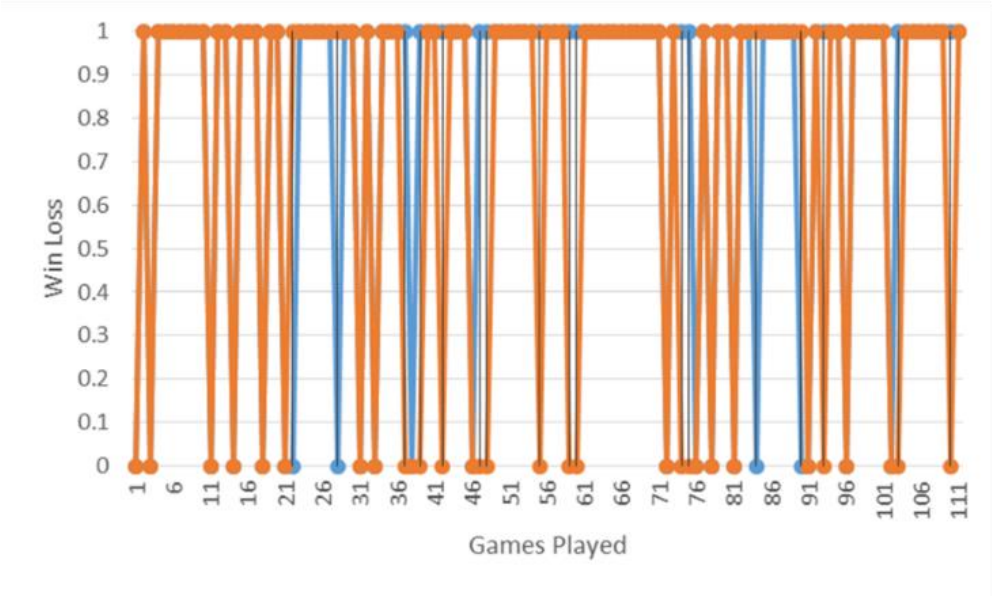


Figure 5. An example of real data collected and indications when "loss of set" was encountered.

A perfect game by a "player" bot would be consistently playing the winning strategy against the "house" bot. For example, if the house bot were consistently playing low to high, the player bot would sort their cards and play them in an offset to the house bot. If the house bot played their lowest card first, the player bot would play their second lowest card, etc. With this scenario the probability of winning a 5-card hand is approximately 66%. This is confirmed semi-analytically as follows:

$$P(W) = P(4W_1) + \binom{4}{3} P(3W_1)P(L_1 + T_1) + \binom{4}{2} P(2W_1)P(2L_1) * P(W_4)$$
$$+ 2\binom{4}{2} P(2W_1)P(L_1)P(T_1)P(T_4) + \binom{4}{2} P(2W_1)P(2T_1) * P(W_4)$$
$$+ \binom{4}{1} P(W_1)P(3T_1) * P(W_4 + T_4) + P(4T_1)P(W_4)$$

X_1 represents Win, Loss, or Tie on cards 1 through 4, and X_4 represents Win, Loss, or Tie on card 5. This calculation estimates a hand-winning probability of 0.66. At present, the data logging is set to play 100 hand epochs and the person advances to the next level if they win 60 of those hands. It may be preferable to have the person play *any* number of hands and have them level up once they win 60 hands (which may occur in fewer or

greater than 100 hands overall). This may facilitate an easier classification of players of various ages and cognitive abilities. An easy feature to extract in this case would be the number of games required to "level up".

With a large enough player base, a classifier should be able to classify a number of profile groups. At the highest level, the ML ANN should be able to easily differentiate between a real person and a bot and then sub-classify the person into various profiles. These may be categorized along profiles such as age and background collected over time with real players, or correlated with self- or independent cognitive assessments. Upon seeing a new player, the ML ANN can generalize to a more sophisticated scoring system is also being developed, assigning greater rewards for narrow victories as opposed to lopsided victories. For each card played within a hand, the score would have calculated as ± (13- (PlayerCardVal- BotCardVal)) depending upon whose card was of greater value. Although this scoring system is statistically more difficult to win against, it provides the opportunity for a person to respond differently after seeing the bot's first card played. Simulated probabilities of playing a winning strategy consistently against the bot are approximately 0.63. This strategy also opens the opportunity to potentially evaluate conservative vs. aggressive winning strategies (e.g. minimize loss vs. maximize gain). This allows us to enter into another area of ML- reinforcement learning (RL)-which is conjectured to be more closely aligned to how people learn and adapt than other areas of ML.

Figure 6 illustrates differences in perfect play using a winning strategy for bot vs. bot, for the two scoring mechanisms above. A histogram of scores per hand for weighted scoring is illustrated in Figure 7.

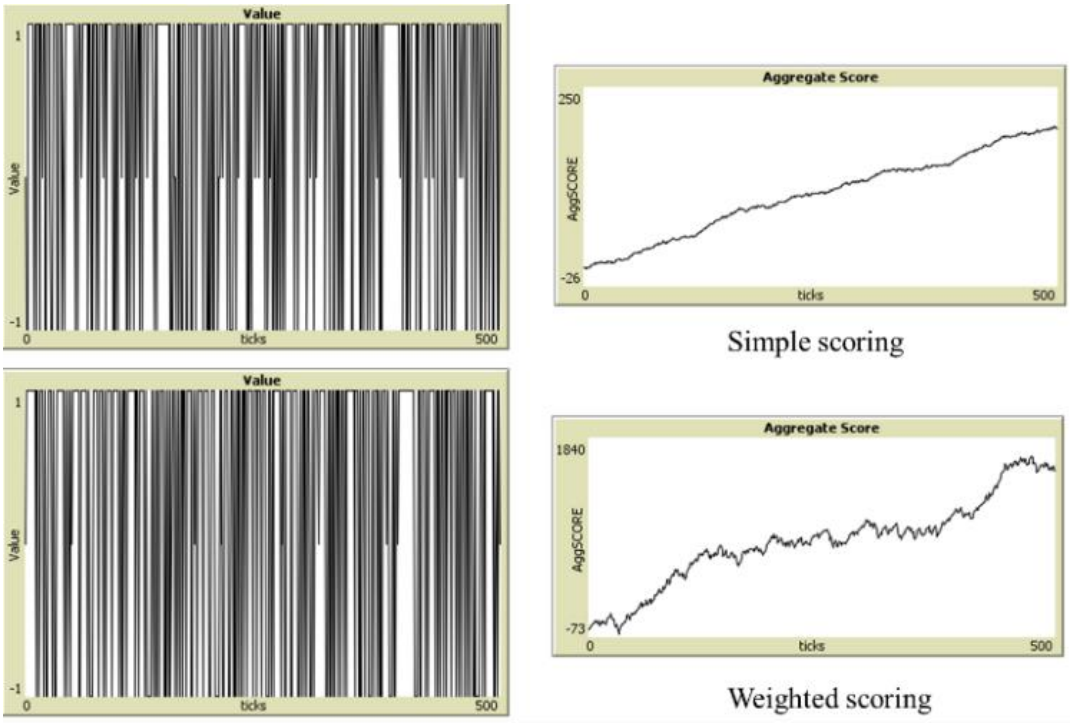


Figure 6. Typical "fingerprints" of Win-Loss Record and Aggregate Scoring from Simple and Weighted Scoring

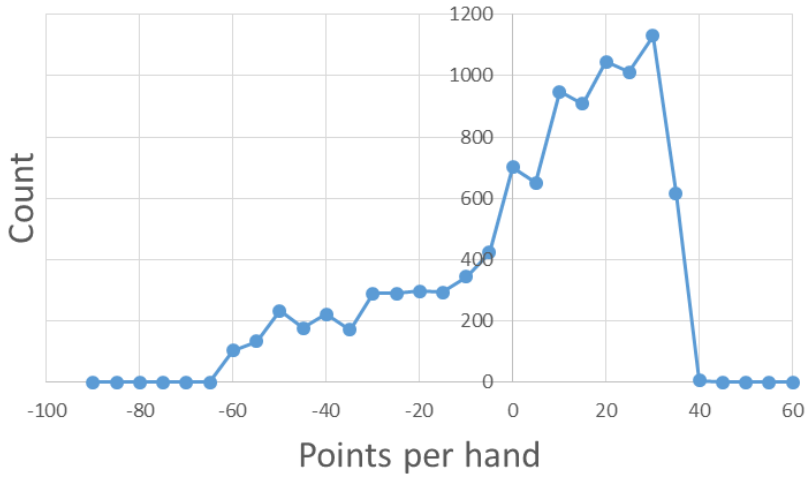


Figure 7. Score per Hand Histogram with Weighted Scoring

The histogram was generated from 10,000 games of bot vs. bot with one bot (the player bot) consistently playing a winning strategy against the house bot. An opportunity lies in a person's ability to learn from the environment and alter their policy to mitigate against large losses. This is analogous to the ML technique associated with

Reinforcement Learning and the cliff-walker problem where a more conservative solution may guard against a large but rare loss on average. The simple scoring vs. the weighted scoring is also interesting from the perspective of how people may feel or perceive their success in the game, as the nature of gains is more sporadic, and this can possibly mislead one into believing in winning and losing streaks. There are also considerably fewer ties in the weighted scoring system.

Table 1. Player records of prototype test volunteers

User	**Level 1**		**Level 2**		**Level 3**	
	win	**loss**	**win**	**loss**	**win**	**loss**
U1	142	53	220	74	292	101
U2	70	30				
U3	85	12	160	34	219	50
U4	80	16				
U5	69	28	208	87		
U6	80	19	162	35	256	51
U7	75	26				
U8	76	23	149	50	213	83

Table 1 illustrates a preliminary pilot for testing an early version of the game using weighted scoring. The purpose of the pilot was simply to gain feedback on game hedonics. It is interesting to note the variability in records: U1 was RDM, the oldest of the group who coincidentally corresponded to taking longer to "level up".

It is notable that the cognitive "fingerprints" of RDM and an age 20-something were considerably different. The question arises, "Do they reflect normal age-related cognitive changes?" Possibly, and that only once sufficient data is collected and analyzed would one be able to tell. The eventual goal would be for ML-aided classification to serve as a giant funnel for people at risk, providing a basis upon which to seek more traditional clinician-mediated assessment and subsequent therapeutic or psychosocial care. Ideally the research will result in reaching a significant number of participants upon which to build baseline cognitive "fingerprints" across demographic profiles from which anomalies can be identified.

WarCAT: Technical considerations

WarCAT has been developed using Unity for the front end, a custom Java middle tier server and a MySQL backend data repository. Some of the technical challenges with developing these types of games is that WarCAT is in constant communication with a server that is logging all player actions. In effect, this makes the game a thin client as it is important that the server have complete knowledge of all hands dealt and how individual cards are played in order to eventually analyze data for subtle changes associated with strategy recognition, recall, loss of set, and perseveration.

An Android version of the game is available and accessible by contacting the corresponding author. As modifications and suggestions are being incorporated in an ongoing development fashion, WarCAT is not available on GooglePlay at this time. An iOS version is also being tested at this time. When stable, both will be released for download.

It is worth noting that WarCAT includes the capability to embed self-report survey questions and embed validated mental health assessment methods such as associated word pairings, to ultimately correlate player statistics with electronic reproductions of more traditional measures of mental health assessment. Although not implemented at this time, aspects of assessment methods such as those found in MoCA and the MMSE can be tailored for a mobile device as well.

Problem Solving: "Lock Picking"

"Lock Picking" Framework section was omitted as it was unrelated to the present work.

Summary and conclusions

The first game presented, WarCAT, is fairly well developed as a mobile game with emphasis on assessment of the cognitive functions of recognition (learning a strategy) and recall (applying that strategy consistently) to the gameplay. The game itself is fairly complex as one of the underlying algorithmic problems is sorting, compounded by requiring card ordering that is not necessarily monotonic. That is, a person may have to order cards medium, high, followed by low. The interesting unexamined conjecture of the game is the very stochastic nature of the game itself may induce a memory slip or sufficient confusion because even when the winning strategy is played, one can still lose if one simply has been dealt a bad hand of cards in relation to the bot's hand. This allows us to build confusion matrices and use methods such as ROC curves to classify player behavior. Additionally, data collected during play can serve as input to a ML algorithm

to facilitate classification, potentially identifying subtle cognitive changes and deterioration.

In addition to the "Lock Picking" and WarCAT games, the opportunities exist to gather other data such as event timing, data from the accelerometer, and data from the front facing camera of the mobile device. These data can provide additional information on how a person is playing a game, including potential difficulties while selecting cards or turning tumblers. The analogy with traditional assessment instruments is that clinicians administering a test may notice, for example, an unusually strong grip on a pencil when performing a drawing a simple shape. In this case, the grip (ancillary to the task) provides as much data to the clinician as the actual shape drawn (the main task). In the case of

mobile devices, eye-tracking technologies using front-facing cameras may provide additional insights into how a score is achieved. Although this work considers only simple mobile games with no additional hardware requirements, it is also worth mentioning the role Brain Computer Interfacing (BCI) technology may play in cognitive function assessment during game playing.

WarCAT and "Lock Picking" are simple exemplars of a genre of games that have the potential to provide the research community with a tool to examine the relationship between cognitive function and aging. It will provide clinicians with an additional tool for identifying potential pre-symptomatic cognitive decline through long term game-interaction data collection, complementing traditional instruments currently in use and potentially allowing measurement of multiple cognitive domains. The mobile game is designed to be an engaging activity for users, and potentially less intimidating than other dementia screening methods typically administered by family physicians and other health care providers. The collected data will offer an opportunity to better understand the relationship between normal memory inefficiency and age, ultimately leading to better statistical models and follow-on insights into pre-symptomatic dementias and other cognitive impairments.

The key technology innovation is in the gamification of cognitive assessment (as opposed to the simple reproduction of traditional assessment tools in electronic or mobile formats), or the application of Smartphone instruments to the assessment of various cognitive functions related to learning, retention, attention, and memory applied to game strategy. These benefits will be achieved through the use of statistical models and more advanced analytic methods (machine learning) as data collection methods and volume improves. Once validated, the work can be integrated into much

larger aging related initiatives and toolsets.

Acknowledgements

Authors Leduc-McNiven, White, Zheng, and McLeod designed and developed the apps. Authors McLeod and Friesen designed the implementation and evaluation strategy. Authors McLeod and Friesen prepared the manuscript. Prototype app development was funded under Natural sciences & Engineering Research Council of Canada grants 397751-12 to author Friesen and 6712-12 to author McLeod.

Conflicts of interest

None declared.

References

1. Petersen, R.D., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.. (1999). Mild cognitive impairment: Clinical Characterization and outcome. Arch Neurol, 56(3), 303–308. doi:10.1001/archneur.56.3.303

2. Jack, C.R., & Holtzman, D.M. (2013). Biomarker modeling of Alzheimer's disease. Neuron, 80(6), 1347-1358. DOI: http://dx.doi.org/10.1016/j.neuron.2013.12.003 PMID: 24360540

3. Spaan, P.E.J. (2016). Episodic and semantic memory impairments in (very) early Alzheimer's disease: The diagnostic accuracy of paired-associate learning formats. Cogent Psychology, 3(1), 1125076. https://doi.org/10.1080/23311908.2015.1125076 (No PMID available)

4. World Health Organization (WHO) and Alzheimer's Disease International, Dementia: A Public Health Priority (Geneva: WHO, 2012). ISBN: 978 92 4 156445 8 (No PMID available)

5. Prince, M., Bryce, R. and Ferri, C., (2011).World Alzheimer Report 2011: The benefits of early diagnosis and intervention. Alzheimer's Disease International. (No PMID available)

6. Tong, T., Chignell, M., Tierney, M.C., & Lee, J. (2016). A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games, 4(1). PMID: 27234145

7. Hagler, S., Jimison, H.B., & Pavel, M. (2014). Assessing executive function using a computer game: Computational modeling of cognitive processes. IEEE Journal of Giomedical and Health Informatics, 18(4), 1442-1452. PMID: 25014944

8. Jimison, H., & Pavel, M. (2006, August). Embedded assessment algorithms

within home-based cognitive computer game exercises for elders. In Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE (pp. 6101-6104). IEEE. DOI: 10.1109/IEMBS.2006.260303

9. Basak, C., Voss, M.W., Erickson, K.I., Boot, W.R., & Kramer, A.F. (2011). Regional differences in brain volume predict the acquisition of skill in a complex real-time strategy videogame. Brain and Cognition, 76(3), 407-414. PMID: 21546146

10. Thompson, O., Barrett, S., Patterson, C., & Craig, D. (2012). Examining the neurocognitive validity of commercially available, smartphone-based puzzle games. Psychology, 3(07), 525. http://dx.doi.org/10.4236/psych.2012.37076

11. Manero, B., Torrente, J., Freire, M., & Fernández-Manjón, B. (2016). An instrument to build a gamer clustering framework according to gaming preferences and habits. Computers in Human Behavior, 62, 353-363. https://doi.org/10.1016/j.chb.2016.03.085

Paper 4: Serious Games and Machine Learning for Detecting Mild Cognitive Impairment

Kyle Leduc-McNiven, Ryan T. Dion, Mahmood Aljumaili, Robert D. McLeod, Marcia R. Friesen
Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, MB, Canada
robert.mcleod@umanitoba.ca

Kai Gutenschwager
Institute of Information Engineering
Ostfalia University of Applied Sciences, Wolfenbüttel, Germany
k.gutenschwager@ostfalia.

***Abstract*—Serious Games combined with Machine Learning (ML) have the potential to provide a complementary means of detecting early signs of mild cognitive impairment (MCI), which is often seen as a precursor to more serious forms of dementia. As with all diseases and illnesses, the ability to mitigate the impact of the illness is directly correlated to early detection and therapeutic intervention. The research described here presents a serious game for mobile devices that captures a "cognitive fingerprint" of a person's play that can subsequently be used as the input to a machine classifier. The objective of the research is to demonstrate that data collected from serious games can be used to train a neural network to classify cognitive difficulties that may be outside the scope of normal age related cognitive decline.**

Keywords—Serious games, mild cognitive impairment, machine learning.

INTRODUCTION

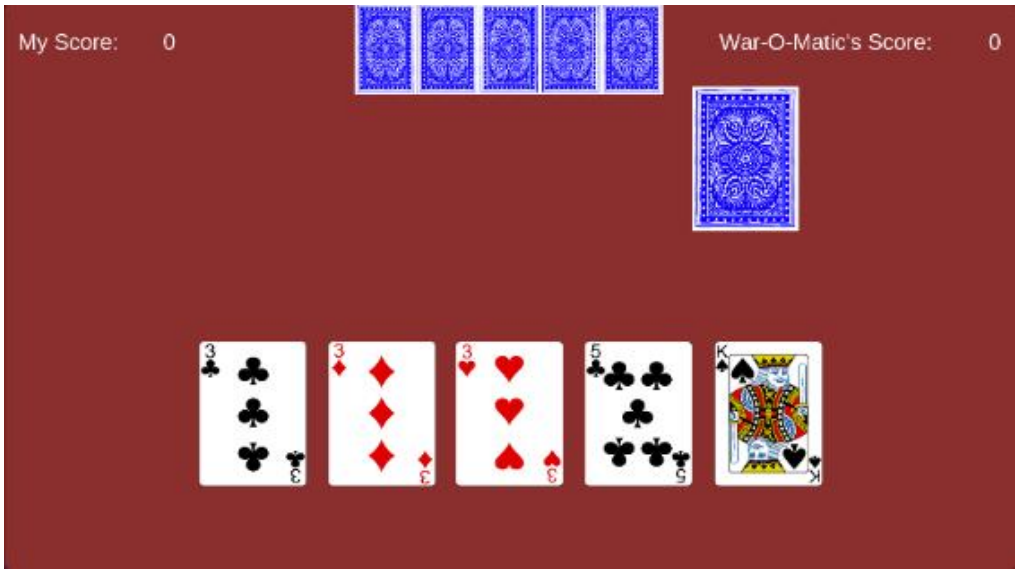
Mild cognitive impairment (MCI) constitutes a clinical entity differentiated from healthy control subjects and those with very mild Alzheimer's disease [1]. It is an intermediate stage between anticipated and normal age-related cognitive decline and the significant impairments associated with dementia. MCI compromises memory, language, thinking and judgment in ways that are more significant than in normal aging [2]. As with most diseases, early recognition of MCI and detecting subtle changes in the brain has the potential to improve the efficacy of therapeutics [3][4]. The role of serious games combined with machine learning (ML) data analysis has the potential to become a technology-mediated assessment tool that is complementary to clinical early diagnosis and traditional assessment tools.

Serious games are typified as being for purposes other than entertainment, although an entertaining serious game has the potential to reach a greater player base than a less engaging game. Most serious games are designed for rehabilitation or possibly engaging and improving functions such as memory and/or recall. Others are being investigated that

integrate a virtual reality component. From our perspective and purposes in developing a game for MCI detection, an ideal serious game would be as engaging as possible, facilitating a large population of users, simple to play on a mobile device, and capable enough to be able to extract detailed data that represents a "cognitive fingerprint" of a player's behavior or play. This data, capturing strategy learning, memory and recall, confusion and strategy recovery, strategy optimization and other high level executive functions can then be classified using artificial neural networks to help extract anomalies from normal age related cognitive decline.

CONTEXT

The serious game developed in the research group is called WarCAT (War Cognitive Assessment Tool). It is important to have an overview of the game as context for the discussion and results that follow. The game is based on the familiar card game of WAR. Our variation is that a hand of 5 cards is dealt to the player, who can play them in any order they choose. The person is playing against a "bot" or machine that has also been dealt 5 cards (unseen by the player). The bot's strategy of play is to play its cards one at a time in descending order. There are other variations of bot strategy, but for the purposes here, this particular strategy is sufficiently representative.



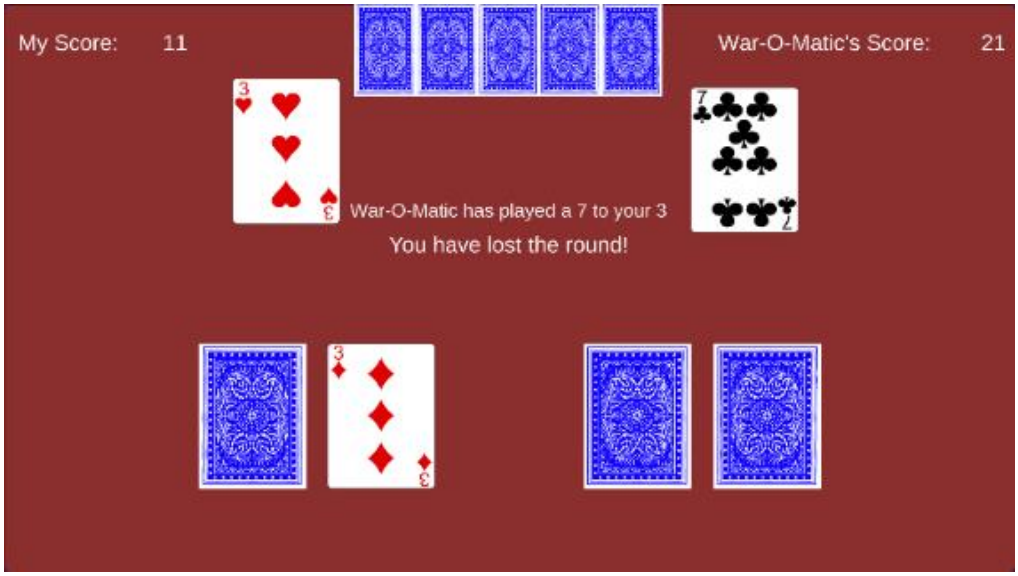


Fig. 1 Screenshot of one hand and instance of play.

Fig. 1 illustrates a hand that was dealt and an instance of play where the player selected a 3 as the fourth card played against the bot's fourth card which was a 7. In this case the bot won the hand. A game lasts for one hundred (settable parameter) hands of play at which time a player levels up. There are a programmable number of levels but for purposes of discussion, three will suffice. In the base version of the game, the bot does not change its strategy at subsequent levels. However, other variations of the game have the both changing strategy at subsequent levels, e.g. playing its cards in ascending order or other variations for the duration of the level.

At this time, as much player data as possible is collected, including the card played for each hand, whether the player won or lost the hand, the score, and timing information. From this, considerable information can be inferred as to whether (and how quickly) the player had recognized and countered the bot's strategy, was remembering the strategy and consistently beating the bot, or was getting confused or suffering from momentary "loss of set" – i.e. forgetting the strategy, and then re-learning it (or not) at some later point in the game. As the cards are dealt stochastically, there is an opportunity for confusion as a consequence of knowing that one played a winning strategy yet still losing the hand. This typically happens 25% of the time with the scoring system currently in use.

The current scoring system scores a narrow margin of victory more significantly than by beating the bot by a wide margin. The score for each card played with a hand is:

$$score = 13 - (Winning\ card\ value - Losing\ card\ value)$$

unless there is a tie, in which case the score for those cards is 0. Alternatively, a point could be awarded to the player or bot depending upon who simply had the higher card.

The difficulty with the latter is that it is more difficult to play a winning strategy and more consistently win. With the one point per card strategy, a person can expect to lose (as a consequence of the stochastic nature of the cards) approximately 35% of the time. It is easier to learn a winning strategy through reinforcement if one wins more often.

As a first instance of analyzing or creating a "cognitive fingerprint," each hand's result can be cast into the entries of a 3x3 confusion matrix as shown below. The nomenclature is as follows: W/W is a Win when you should have Won based on the cards dealt. L/L is a Loss when you should have Lost based on the cards dealt; T/T is a Tie when you should have Tied, based on the cards dealt; L/W is a Loss when you should have Won based on the cards dealt; and so forth.

TABLE I: Confusion Matrix

W/W	W/L	W/T
L/W	L/L	L/T
T/W	T/L	T/T

Since each hand played provides an entry in the confusion matrix, inputs or features are generated that can be used as inputs into neural networks for classification. In this manner, a feature vector is created of the length equal to the number of hands played. For example, if a person played their winning strategy for 100 hands, a vector of 100 elements of {W/W, L/L, T/T} would be generated. As a person deviated from their winning strategy, any element from the confusion matrix could occur. These vectors, although difficult for a person to interpret, are ideal for many types of ML methods.

Preliminary Machine Learning

The initial foray into ML was to determine the approximate number of players or amount of data that would be required for a machine to be able to classify play. To accomplish this, synthetic data was generated from bots playing bots. From this, up to 110,000 bots played with various degrees of impairment from perfect play. Bot impairment was instantiated simply as playing randomly for that hand (Bernoulli trials). Bot levels of impairment ranged from 0% to 100% impaired at 25% increments. Three levels of impairment are illustrated in Fig. 2.

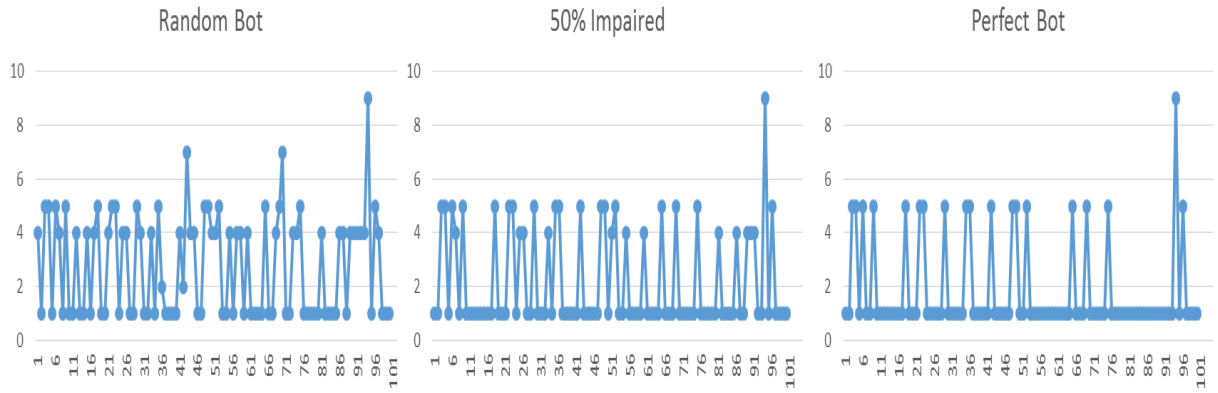


Fig. 2 Three "fingerprints" of play for bots playing 100% randomly, 50% randomly, and perfectly.

For preliminary ML training and testing purposes, we used a densely connected neural network with 103 inputs which included the features generated from confusion matrix (e.g. Fig. 2), as well as aggregate wins, losses and ties.

Within this configuration, 110,000 games were played by bots with five classification labels: Random Play (100% Impairment), 75% Impairment, 50% Impairment, 25% Impairment, and Perfect Play (0% Impairment). The neural network used was a dense neural network (DNN) with five output neurons, and internally two hidden layers of 150 neurons each. The training used stochastic gradient descent, trained on 100,000 training patterns, and tested using 10,000 patterns. The TensorFlow framework and high level APIs were used.

Table II summarizes the training and testing accuracy of the neural network. These are very respectable values for accuracy and do not provide any evidence for overfitting or biases. It should also be noted that that all the classification errors were of the closest match. For example, if the classification was Random Play, the error - if made - was a prediction of 75% Impaired. These are errors of the best kind in that they are close to the correct classification.

TABLE II: Training and test accuracy after learning using the "cognitive fingerprint" data

Epoch	Training Accuracy	Testing Accuracy
500	85.6	82.8
1000	92.0	87.8
1500	94.6	93.1
2000	97.5	94.2
2500	98.4	96.2

While this model was run using hand-by-hand data, a similar model was run just using aggregate Win/Loss/Tie data (number of wins, losses and ties) as a sanity check. Aggregate data is typically descriptive and easily interpreted by people, however the subtleties of the actual data are lost. At best, the DNN was only 60% accurate in training and prediction with aggregate data.

Improving the Synthetic Data

Although encouraging preliminary ML accuracies were achieved, the error model used is not representative of how people tend to become confused and then subsequently recover and return to play their winning strategy they had learned. As a reasonable first extension, a bursty error model was introduced. This model was inspired by early models used for error correction associated with transmitting data over noisy channels. In these cases, the disturbance in the channel extends over many bits or symbols.

The premise here is that a person may undergo a degree of confusion that is dependent upon previous hands played. The previous error model used was simply based on independent Bernoulli trials, similar to that produced by flipping a coin. In this case, a head would represent playing the correct strategy whereas a tail would be a random strategy. As a first attempt to more realistically model play and degrees of impairment, a simple Gilbert-Elliott two state Markov process [5] for generating errors is considered.

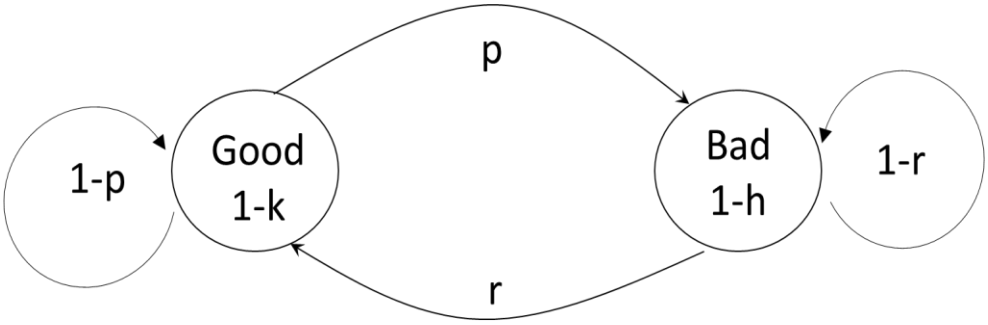


Fig. 2: Gilbert-Elliott model generating a 2-state Markov modulated error process

Errors are generated at rates of 1-k while in the Good state and 1-h while in the Bad state. The steady state error is given by:

$$p_e = (1-k)\frac{r}{r+p} + (1-h)\frac{p}{r+p}$$

An illustration of the burstyness of the error patterns generated is shown below:

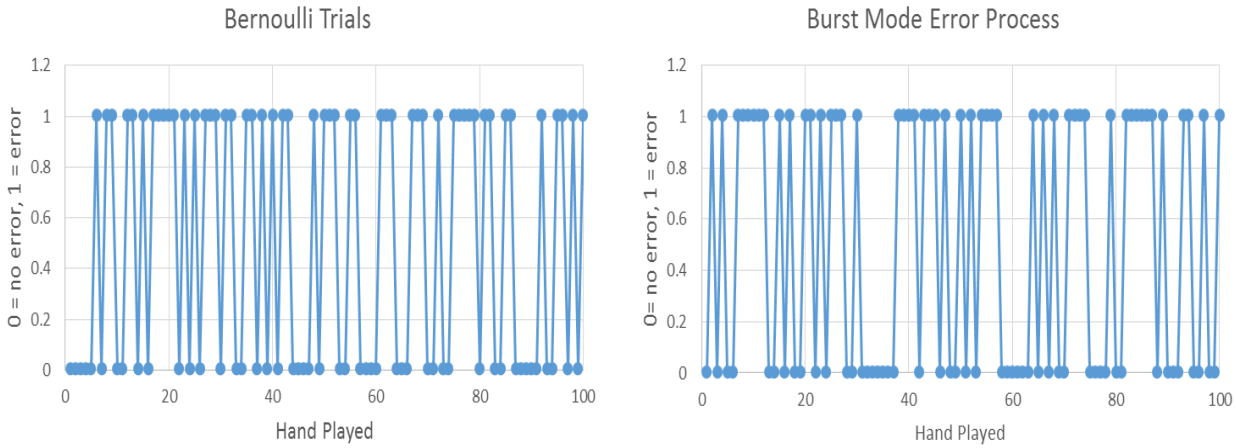


Fig. 3: Burstyness of the Bernoulli and Burst Mode modulated error processes (p=0.25, r=0.25, k=0.9, h=0.1).

In Fig. 3, the Bernoulli trials were generated using a fair coin with p=0.5, The Burst Mode "loss of set" or error in play was generated with parameters p=0.25, r=0.25, k=0.9, and h=0.1. The number of errors generated was 51 and 50 out of 100 for the Burst Mode and Bernoulli trials error generating modes of play respectively.

Fig. 4 illustrates the distribution of the error process. The mean for the Burst Mode was 51.4 (SD=6.6) and for the Bernoulli trials was 51.1 (SD=4.9), averaged over 1000 games (100,000 hands). This would imply that if this is a better model for "loss of set", the burst mode would likely be harder to classify. A potential advantage is that the feature set may be richer.

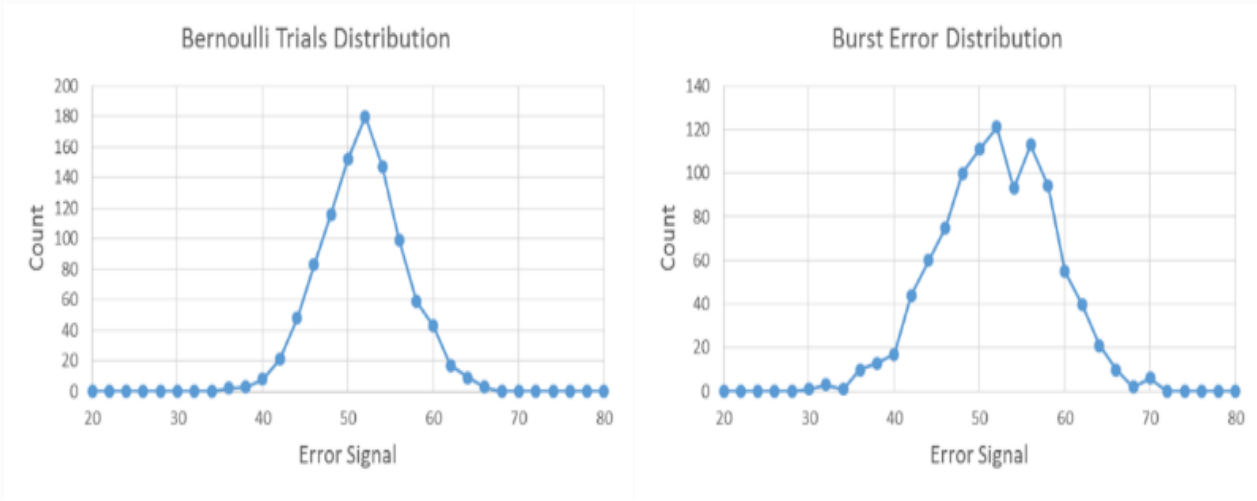


Fig. 4 Frequency distributions of the error process (p=0.25, r=0.25, k=0.9, h=0.1).

Another example with a more bursty flavor is shown in Fig. 5. The Bernoulli trials were generated using a fair coin with p=0.5, The Burst Mode "loss of set" was generated with parameters p=0.2, r=0.2, k=0.0, and h=1.0. The number of errors generated in both cases was 50 and 51 out of 100 for the Burst Mode and Bernoulli trials respectively.

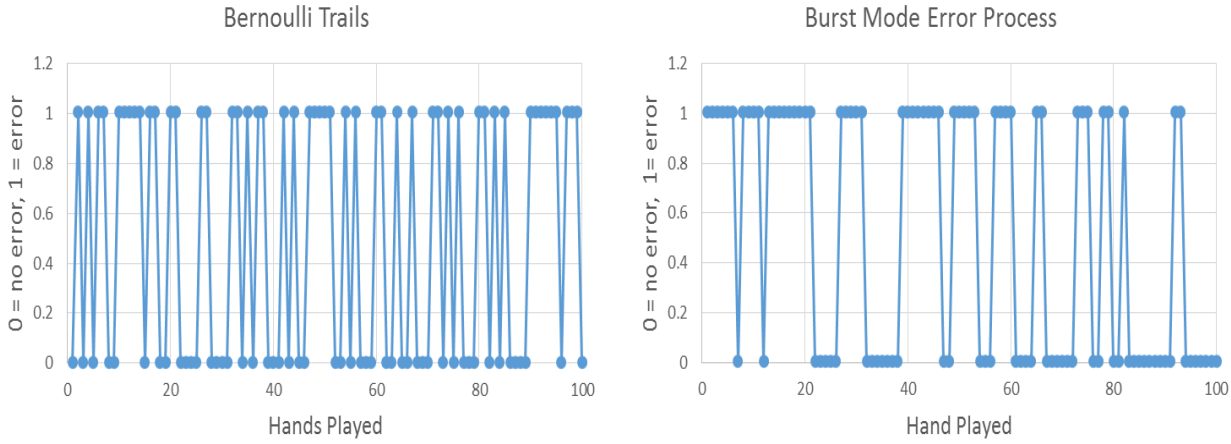


Fig. 5: Burstyness of the Bernoulli and Burst Mode modulated error processes (p=0.2, r=0.2, k=0.0, h=1.0).

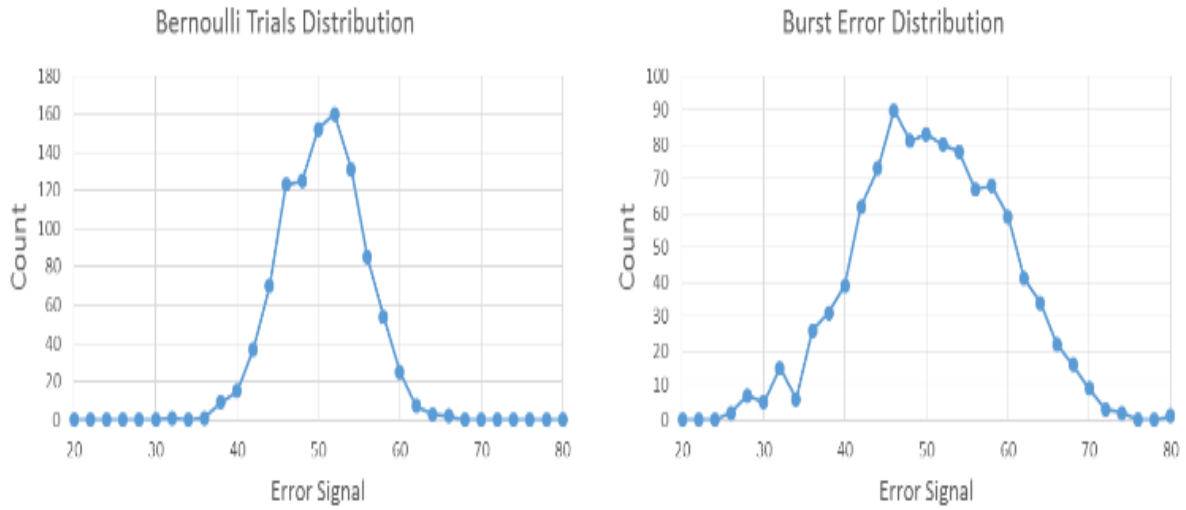


Fig. 6: Frequency distributions of the error process (p=0.2, r=0.2, k=0.0, h=1.0).

Fig. 6 illustrates the distribution of the error process. The mean for the Burst Mode was 50.2 (SD=8.9) and for the Bernoulli trials was 50 (SD=4.9) averaged over 1000 games (100,000 hands). This would imply that if this more bursty model is a better model for

"loss of set", the Burst mode would likely be even harder to classify due to its increased variance.

Values picked to emulate the defective bots are as follows: (These most closely resemble those of Fig. 5.)

TABLE III: Values of prhk used to infer various levels of impairment

Defect Level	p r h k	Average % Win	Stdev
Perfect Play	N/A	74	4.3
25% Defect Play	p=0.13333; r=0.43; h=0.0; k = 1.0;	68	4.81
50% Defect Play	p=0.2;r=0.2;k=1.0; h=0.0;	61	5.4
75% Defect Play	p=0.415; r=0.133333; h=0.0; k = 1.0;	55	5.16
100% Defect Play	N/A	49	5

The Average is the average number of wins in 10,000 hands and the Stdev is the standard deviation associated with the distribution of wins in 10,000 hands. This is narrower than the error signal distribution as the error signal simply indicates when a bot is not playing the winning strategy. In that case, the bot may still win by virtue of the cards dealt, and the same applies to the bot playing the losing strategy (i.e. the bot may still win).

Table IV summarizes the training and testing accuracy of the DNN neural network when using the burst error mode. Again, these are very respectable values for training accuracy; however, the testing accuracy is unexpectedly and disappointedly poor.

TABLE IV: Training and test accuracy after learning using the "cognitive fingerprint" data and burst mode error signal

Epoch	Training Accuracy	Testing Accuracy
500	84.49	67.44
1000	91.5	66.6
1500	95.0	65.0
2000	NA	NA
2500	NA	NA

A surprising result followed: Another dataset was generated that simply had inputs for whether the player bot was playing the correct strategy or simply playing randomly. This produces a simplified cognitive fingerprint as opposed to the ones generated from the full confusion matrix. Again, this was for the games played that used the burst mode errors.

As illustrated in Table V, the training and testing accuracy was quite reasonable. Even after 200 epochs, none were misclassified by more than one category. Possible slight overfitting is observed just beyond 300 trailing epochs.

TABLE V: Training and test accuracy after learning using the simplified "cognitive fingerprint" data and burst mode error signal

Epoch	Training Accuracy	Testing Accuracy
100	95.5	95.0
200	96.1	95.2
300	96.7	95.1
400	97.3	94.8
500	97.9	94.5
1500	99.9	93.01

Preliminary classification results using a more sophisticated neural network (CNN).

As we are interpreting the ML input vectors as a type of one-dimensional image, the ML progression is towards that of a convolutional neural network. As a template and initial CNN instance, the CNN MNIST classifier[1] was used. The rationale being that the complexity of the MNIST data set was assumed to be comparable to the complexity of the "cognitive fingerprints". A difference is that the MNIST data is an actual image 28 by 28 pixels whereas the "cognitive fingerprints" are one dimensional [100 by 1]. Our conjecture with the "cognitive fingerprint" is that there is a relationship between adjacent hands of play modeled with the burst random signal generator. As such, the kernel size for the first convolutional layer was set to [1,50] with the number of filters set to 32. This was followed by a pooling layer with pool_size=[1, 2] and strides=2. The second convolutional layer had the kernel size set to [1,10] with the number of filters set to 64, followed by a similar pooling layer, followed by a similar dense neural network as the MNIST classifier and logits output layer with our 5 output labels.

TABLE VI: CNN Training and test accuracy after learning using the simplified "cognitive fingerprint" data and burst mode error signal

Epoch	Training Accuracy	Testing Accuracy
100	93.52	93.17
200	94.38	94.09
300	94.66	94.52
400	95.06	94.83
500	95.26	95.01
1500	96.26	95.24

[1] https://www.tensorflow.org/tutorials/estimators/cnn

Data used was inferred “perfect or not” play (as opposed the full Confusion Matrix data) with the burst mode error model. Preliminary results are shown in Table VI. These are encouraging results tempered with an understanding that the CNN was preliminary and not tuned and the data used synthetic, albeit with a burst mode loss of set modulating the level of impairment. Effectively, these results are a consequence of an accurate inference of whether the play was impaired or not. Having an accurate inference of whether the play was impaired or not would lead to similar (a few % less accurate) classification from simple statistics alone.

Re-interpreting Strategy

Much of the following is based on actual play and observations made during play, which has led to re-examining a person’s optimal strategy. A reasonably good strategy which we could already consider as “perfect play” is to consistently play the lowest card first followed by highest to lowest. In this strategy, the first card is sacrificed to see what the bot’s highest card is (where the bot’s strategy has been identified to play its cards in descending order). This is a winning strategy leading to wins approximately 75% of the time, but doesn’t account for more subtle decision making during the hand. However, there are other strategies which may also lead to good or even better results. In order to evaluate the 120 possible basic strategies, defined as the possible permutations of the five cards, we performed a simulation experiment where 10M games have been played against a bot that always plays its cards high to low. We further analyzed the results by storing information about the lowest and the highest card value, as more advanced players might use this information as input to a more adaptive strategy. Our results show that looking at all 78 possible combinations of lowest and highest card, 23 different permutations of cards lead to the best results at an average for at least one combination of lowest and highest card.

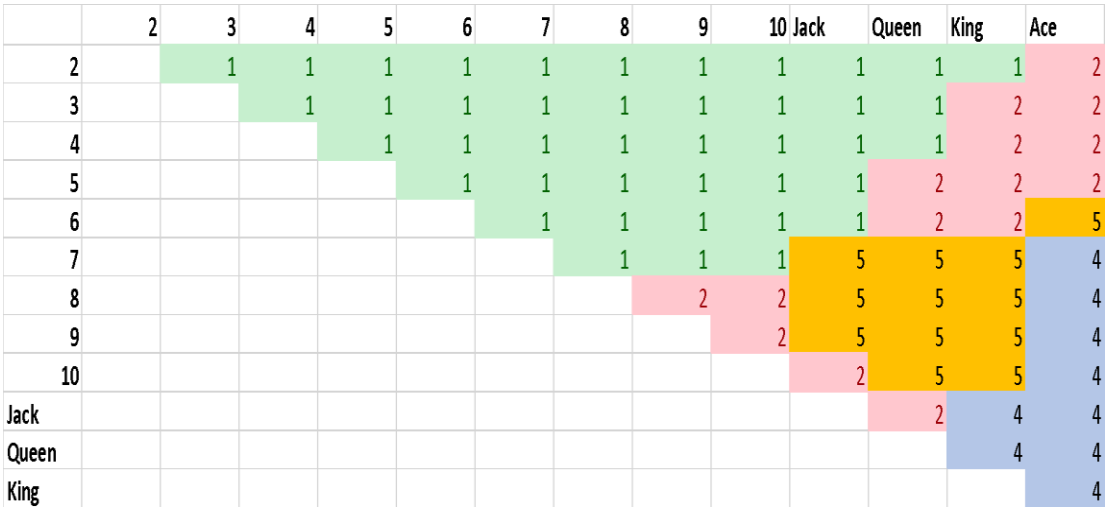
If a player always decides on this adaptive strategy, the respective sequence to play their cards based on the present combination of the lowest and highest card (in their hand), the percentage of wins increases to about 83.4%. The main strategies are given in Table VI.

	2	3	4	5	6	7	8	9	10	Jack	Queen	King	Ace
2		18	9	13	13	9	1	1	1	1	1	1	21
3			23	9	12	6	6	1	1	1	1	2	2
4				12	22	6	6	1	1	1	1	2	2
5					6	1	1	1	1	3	2	2	2
6						3	1	1	3	3	2	2	5
7							14	3	3	10	5	5	4
8								15	19	10	5	8	4
9									2	5	8	8	4
10										20	17	7	4
Jack											11	7	4
Queen												7	11
King													16

Fig. 7: Winning strategy depending on the lowest and highest card

TABLE VI: Winning strategies with respect to the lowest and highest card. Note that other strategies are minor variations of these strategies

ID	Sequence	Description
1	**1**-**2**-5-4-3	Sacrifice first two lowest cards at first and second position.
2	**1**-5-4-3-2	Sacrifice lowest card at the first position.
3	**1**-5-**2**-4-3	Sacrifice first two lowest cards, but the second best card at third position.
4	5-4-3-2-1	Play high to low
5	5-**1**-4-3-2	Sacrifice lowest card at the second pos.
6	**2**-**1**-5-4-3	Sacrifice first two cards (second best first)
7	5-4-3-**1**-2	Sacrifice lowest card at the forth pos.
8	5-4-**1**-3-2	Sacrifice the lowest card at the third pos.



	2	3	4	5	6	7	8	9	10	Jack	Queen	King	Ace
2		1	1	1	1	1	1	1	1	1	1	1	2
3			1	1	1	1	1	1	1	1	1	2	2
4				1	1	1	1	1	1	1	1	2	2
5					1	1	1	1	1	1	2	2	2
6						1	1	1	1	1	2	2	5
7							1	1	1	5	5	5	4
8								2	2	5	5	5	4
9									2	5	5	5	4
10										2	5	5	4
Jack											2	4	4
Queen												4	4
King													4

Fig. 8: Winning strategy depending on the lowest and highest card (simplified)

If we reduce the number of different strategies (IDs 1, 2, 4, 5) of play to four we obtain a rather manageable set of strategies and combinations of lowest and highest cards and when to use them. The percentage of games won is still very high with a value of about 82.7%.

Note that if a player also takes into account the first card the bot plays and adapts their strategy based on that card, the chance of winning further increases by about one percentage point.

Looking at the limited number of games a player completes in one round, the difference between these two approaches is given in Fig. 7 and 8 and is hardly recognizable and not of statistical significance. We therefore differentiate a few main strategies that actually lead to statistically significantly different results that a player may actually recognize playing only a few rounds.

Table VII provides an estimate of the expected points awarded based on the strategy played. For example, if a player played at random, the average number of points expected would be 0. If they developed or learned a strategy of consistently sacrificing their first card followed by playing their cards from high to low, their expected points per hand would be 16.6. Within one level, a person would play 50 hands and if they proceeded to play that strategy they would end up with approximately 50*16.1 points.

Overall, we can vary the strategy a player follows as the first step, and then analyze if the player gets distracted from that strategy over time. However, a player may tend to choose a different strategy willingly over time to improve their overall score. These variations make it more difficult to infer moments of confusion or loss of set.

TABLE VII: Expected points awarded per hand, based on strategy played

ID	Strategy	Avg. points scored
1	Random	0
2	High to low	0
3	Low to high	0
4	Sacrifice first card, rest high to low	16.6
5	Sacrifice first two lowest cards, rest high to low (lowest at pos. 1)	16.1
6	Adaptive (1): If the value of second lowest card < 8 then sacrifice two cards, else one	18.1
7	Adaptive (2)	20

The first three strategies can be differentiated by the sequence played, even though the average score does not differ. We use a simple distance measure for comparing two sequences played, denoted as P1 and P2, where pos_j states the position of the j-lowest card in the sequence:

$$dist = \sum_{j=1}^{5} abs(pos_j^{P1} - pos_j^{P2})$$

Computing the average distance for the last n games to the (best) sequence of the respective main strategy (2-7), we can analyze the current distraction as a deviation from the optimal sequence of the strategy the player follows momentarily (except for random play). Taking into account the average score obtained using a specific strategy, we can further compare the average score the player achieved with the respective expected value. This information might give further insights into why a player changes a strategy, i.e., if the actual score is lower than the expected value (computed by our simulation experiments), there might be a higher probability that a player gets distracted from the strategy currently followed.

The classification on whether and how significantly a player's behavior is related to a cognitive decline needs to be corrected by this influence. We furthermore need to take into account a learning curve, such that we compare a player's current play to the best strategy learned so far. This means that the "perfect play", as introduced in Sec. III, is actually to be considered as dynamic if we want to judge the influence of cognitive decline.

REINFORCEMENT LEARNING

The problem is now more readily identified as having labeled data that actually reflects various levels of impaired play. The approach that is now being undertaken takes its lead from AI techniques being applied to game agents or bots [6]. In the case of AI agents in games, typically the objective is not to outsmart players. The objective is to have the

agents be as smart as necessary to provide fun and engagement. The objective is not to push the limits of ML within the bot, as would be the case in many ML optimization instances. For our purposes the ML bot needs to be imperfect, imitating a human-like behavior. With this objective the goal is to have bots train for various periods of time. Bots trained with minimal iterations will essentially have no clues on what to do, whatsoever. While those that have been trained longer will have learned strategies that enable them to win consistently. As such, bots with various levels of training will represent human play with inverse various levels of impairment.

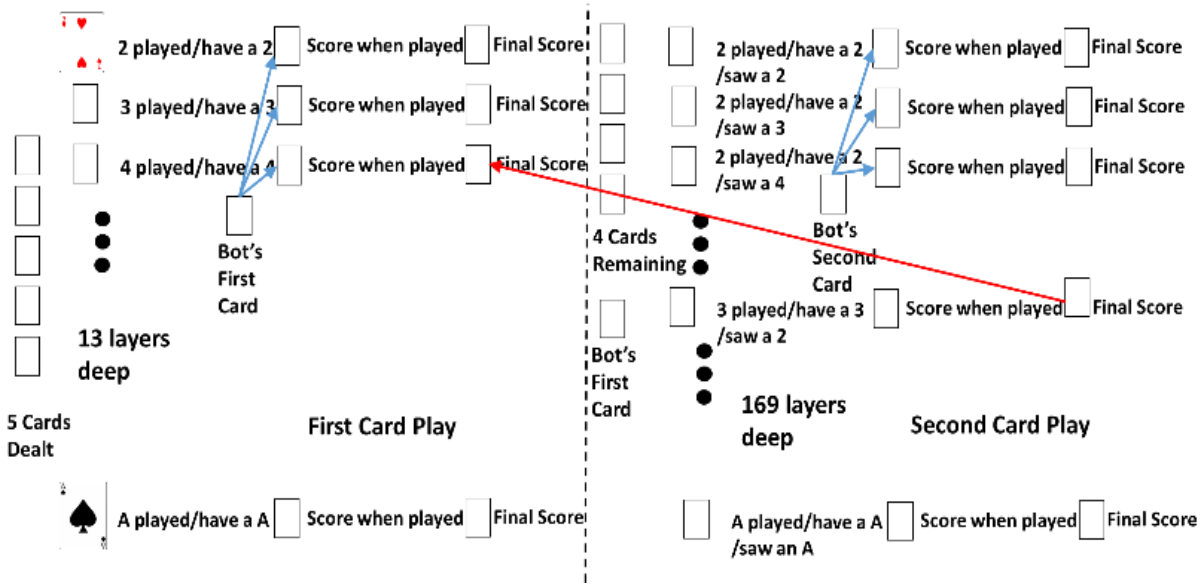


Fig. 9: First two stages for learning human like strategies

An illustration of the proposed strategy learning method is shown in Fig. 9. Figure 9 represents all combinations of conditional play with the first two cards played. Strategies of play will be learned by playing against a bot that is dealt random 5 card hands where the bot plays its cards high to low. Scoring boxes are shown that are populated during training. The immediate score when each card is played as well as the total score for that hand upon completion of the hand. Effectively this process resembles that of finding the most suitable paths through a large decision network. This finding of suitable paths or culling poor paths is also illustrated in Fig. 10. In a simplified scenario the graph would in effect would initially be a complete network of 5 layers of 12,144,144,144, 12 nodes respectively. Each Layer corresponds to the card to be played, i.e. the layers in the middle are not conventional hidden layers.

With a simple reinforcement learning algorithms preferable trajectories will emerge. During training cards will initially be played at random and scores tabulated. After a series of games preferable strategies will become more heavily weighted. All moves are

conditional. For this initial foray into using reinforcement learning only the last card played by the bot will be used as a condition. After sufficient training the more favorable trajectories through the game can be used to accelerate the reinforcement. This phase driven by learned pdf's will be analogous to a self-organizing network.

Figure 10 illustrates the original network of strategies (left-hand figure, in which all sequences are equally likely) and the network of strategies that emerge as being more favorable (right-hand figure). The clustering in the left-hand side of the trained network (right-hand figure) illustrate the initial sacrificing of a low card or the playing of a high card, if held.

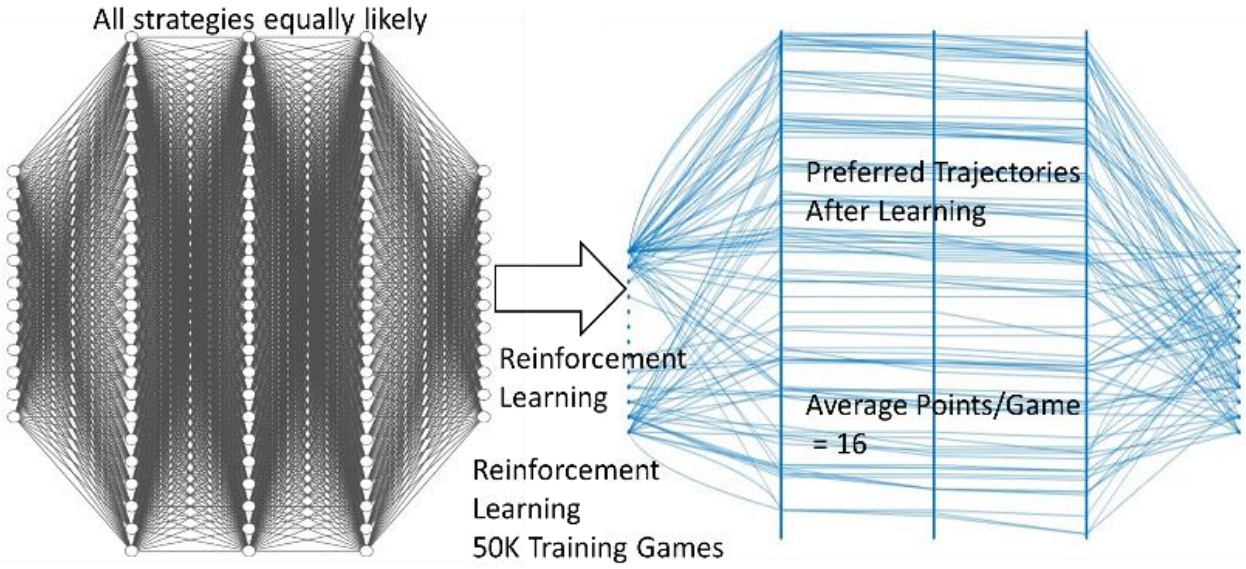


Fig. 10: Emergent human like strategies[2]

DISCUSSION

This version of WarCAT consists of three "levels", demarcated by an overall score and win/loss record where the bot continues to play a consistent strategy through each level. The thought is that a person may learn the mixed strategies as a means to optimize their score. It is important that hand by hand data be used to construct input for ML techniques such as DNN or CNN, as simply using aggregate overall scores would occlude momentary difficulties a person was having with recall or reasoning. At this time efforts are being made to improve the interpretation of play and the subsequent interpretation of distracted play and inference of cognitive decline.

Even for this relatively simple game, although collecting extensive data on how a person played is technically feasible, classifying play in terms of cognitive decline is very

[2] Drawing courtesy http://alexlenail.me/NN-SVG/LeNet.html

complicated. In light of not have extensive real player data, synthetic data and a burst mode model of impairment demonstrated that machine learning may be a viable technique to classify degrees of impairment. Demonstration of play can be found on YouTube at https://www.youtube.com/watch?v=CN-_zz-oIpU&t=30s.

Also, if anyone were interested in playing WarCAT, an Android version (APK) can be obtained from the link in the description box of the video or by contacting the authors.

Acknowledgment

The authors thank Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Archives of neurology. 1999 Mar 1;56(3):303-8..

[2] http://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/home/ovc-20206082

[3] Jack Jr CR, Holtzman DM. Biomarker modeling of Alzheimer's disease. Neuron. 2013 Dec 18;80(6):1347-58.

[4] Spaan PE. Episodic and semantic memory impairments in (very) early Alzheimer's disease: The diagnostic accuracy of paired-associate learning formats. Cogent Psychology. 2016 Dec 31;3(1):1125076.

[5] Elliott EO. Estimates of error rates for codes on burst-noise channels. The Bell System Technical Journal. 1963 Sep;42(5):1977-97.

[6] Snoswell AJ, Snoswell CL. On the use of reinforcement learning for testing game mechanics. Computers in Entertainment. 2018 Apr 27.

Postface

In this postface I will detail my contributions to the papers.

General Contributions

My first and most significant contribution was creating the first working version of the game WarCAT and WarDOG. The game WarCAT is a retooled version of WarDOG made to remove all multiplayer functionality. I was responsible for updating an original framework called MOBRO WAR including its servers and database architecture. I altered MOBROWARS deck of cards used so that it would function like an actual deck of cards with different suits. This was done with switch case statements and nested while loops to make sure all the suits were evenly represented, the odds of a card getting pulled was equal at the start after which the odds would be updated appropriately. The game was designed so that it would be very easy to change anything. The most challenging issue when changing something was whether or not the new information would be stored in the database, or whether the database would need to be updated. This was addressed by adding columns to the database for the new data and adding lines in the server to receive and interpret the incoming game data to store it in the appropriate place. This was done so that regardless of future directions, I and future developers can easily change it to their liking or tailor it to their research question. I created the database tables and chose the information that would be stored to best get and store game data. Details are in the database section explaining exactly how it functions. I made the decision to have the first representation of a player's strategic fingerprint be represented by a bar graph which showed the average card, then used first, second, third, fourth, and fifth cards with cards being represented by their value in relation to the rest of the hand rather than their actual value. An example would be in the hand King, Queen, Jack, 10, 9. The king is the highest value card in the hand and is given the ID 5 to represent that, and the 9 being the lowest value card in the hand would be represented by a 1 ID. This was done so that strategy could be more easily broken down. Since the idea of an adaptive strategy was still to be determined, the assumption was that the optimal strategy of play was to sacrifice the lowest card then play highest to lowest, or play 1-5-4-3-2.

I was the team lead when the first discussions of using machine learning for classification was discussed. I was responsible for the idea of creating synthetic data and created the first simple AI that played the game with varying levels of forgetfulness. There were four levels of play that I designed. Each level had the BOT opponent playing a different strategy so that the player would be forced to come up with a new strategy. The

fourth level of strategy was not discussed in any of the papers but it was meant to be a fun little reward where the game changed from developing a strategy against a particular type of play, to a game where you had to figure out what style of play the opponent was using and then remember the appropriate strategy to use. The expectation was that the strategies developed during the earlier levels would be remembered and used here. The last general contribution was that I helped edit the papers during the editing process.

Paper 1 Contributions

For this paper, I developed a quick version of the game where different hand sizes were tested to see what would allow for the best results for detecting MCI by allowing for forgetfulness. Five cards were determined to be the optimal amount, however the version of the game for multiple hand sizes would be simple to implement due to the nature of the framework created. The only issue would be the database game tables where it looks for five-card hands, but the solutions to that are many and simple. I also looked into the use of eye tracking as mentioned in Chapter 4 in the peripherals section.

Paper 2 Contributions

The major contribution I had to this paper was assisting with a new strategy decision table and determining how to store it correctly such that each game in the database had a corresponding strategy value attached to it. This worked by having a null slot in each game that would be later fed through code that would interpret the cards played and the order they were played, and return a value of either random play, normal strategy, or advanced strategy. Those results were then taken and put into a graphic wheel for visualization which could be easily used by a CNN for classification.

Paper 3 Contributions

For paper three, I contributed in two major ways. First, I was responsible for the prototype for the game levels, where they would function in sets of 100 to facilitate easy computation of percentages and was of sufficient length so that a person could create and forget a strategy. The idea behind the levels was that if the player did not get at least one standard deviation from the norm in wins, then they could not progress and would have to play the 100 games again. The games themselves were very fast paced if you knew what you were doing, 100 games could be played in about 50 minutes or less with the average game length being around 20-30 seconds. The length of games required was later shortened to 50 games. Second, I implemented the games' scoring method in a switch, so that one could switch between the two different scoring methods. If one wanted to go

back to a ‘win equals 1-point’ system, it would be easy to do so; however, the system for scoring seemed to create interesting gameplay and allow for varying strategy so there was never really any need to.

Paper 4 Contributions

I developed the idea for synthetic data and created the first BOT that creates synthetic data which has a loss and forgetfulness metric. The BOT was created so that it would actually play the game so that the proper randomness of the cards could be represented, rather than just an automated program that created server statements. The BOT also had a forgetfulness factor where it would occasionally just play the cards in its hand from left to right. I created the database tables that handled the data that dealt with the creation of strategy allocation. The tables were altered to the specifications given to me by Dr. Gutenschwager and my own judgment. I also updated the app and features that were not fully functional.

Chapter 6: Future Works & Conclusions

Future Work

Clean up game images and animations

The first priority before deployment of the game is to polish the game's aesthetic and its animations. The game itself functions; however, just like in all things, presentation matters greatly. As it is now, the game needs polish in all the panels. The polish should include changing how the buttons look, and making the background of every panel more pleasing to the eye. Other aesthetic changes should be limited to the font and the cards. The animations are mostly fine as is, since the longer the animation, the more it can disrupt the length of play which is important for detecting MCI. However, the animations do need to be altered such that players have longer to view what cards were played in a particular battle and what the results were, as when showing off the game one of the things that was commonly noted was that players were unsure what happened after a card was played and how well they did. This objective requires professionals in art and animation to look through the game and provide feedback.

Machine Learning Methods

One of the biggest problems with using serious games and machine learning (ML) for classification of MCI is that if one lacks properly labeled data, one will end up with overfitting and erroneous classification. This could lead to detecting something that does not exist (false positives) or missing something that does exist (false negatives), both of which have costs. To get around this problem, we attempted to use synthetic data to augment the amount of MCI data available. This leads to several possible solutions: collect a small amount of real data and extend it using ML, or use what data exists from MCI patients and use feature selection methods to narrow down important data points for classification. To use ML to expand the data, there are several different methods. The first method mentioned in Chapter 5 (Paper 4) is to utilize Reinforcement Learning (RL) to inherently create new data for training the classifier. The second possibility is to utilize Generative Adversarial Networks (GANs). GANs are especially useful in this case as it trains not only a generator for the data, but it would also train the classifier at the same time. It does this by having two Neural Networks compete against each other, where the generator is creating new data in an attempt to fool the discriminator into thinking the data is real, and the discriminator is trying to compare

the real data with the fake data so that only real data are detected. The added benefit to this is that since the classification is binary where you either have or do not have MCI, the discriminator essentially learns whether the data fed to it match up with its definition of MCI or not. The drawbacks to using a GAN is that if the original data set for the discriminator to compare against is not large enough, then you may end up with synthesized data that is overfit as well. RL methods do something similar to GANs; however, they feed newly-generated data back into the network if it is deemed as good. Combining a GAN and RL network may be an approach to use a small set of data and then extend it and train the network on MCI detection.

Data gathering from older adults

The next stages for this research project would be to gather the data of people with MCI. MCI is classified as cognitive impairment greater than expected for age and education. By this definition, older adults with memory lapses may not, simply by virtue of age-related memory lapses, be identified as having MCI; however, these data would be useful for creating baselines based on age. This would allow for younger groups to have their data classified based on what age the classifier thinks they are, which could trigger clinical assessment where anomalies are detected. There are also ethical and clinical issues that would be difficult to address, and which are beyond the scope of the research question posited here.

Deployment of the game

The final future step would be to actually deploy the game for large-scale data collection across the population and set up a dedicated server and database. To do this, one would need to set up a proper server rather than running it off of the development machine, and the game would also need the polish described earlier. The difficulty is that successful deployment and uptake depends on player engagement and the appeal of the game's look, feel, and functionality, which is a very difficult problem faced by all game developers, and more so by those attempting to deploy serious games.

Conclusion

The literature supports the exploration of serious games as a viable method for detecting early MCI. The purpose of WarCAT is to offer an additional tool by which to detect cognitive performance over time, such that its players can gain knowledge and seek professional help for a proper diagnosis. With the steps outlined in this thesis, one can reasonably be expected to produce good classification results. Utilizing the varying

machine learning methods and some feature selection, high accuracy classification should be possible.