Quantum Computing: Evaluating Potential Quantification of Projective Psychological Test Scoring

Dan M. Davis
Univ. of Southern California
Long Beach, California
dmdavis@acm.org

Sanad Z. Rizvi & Kenneth Shaw
Univ. of Southern California
Los Angeles, California
mohdsanadzakirizvi@gmail.com
& kshaw@gatech.edu

Mark C. Davis
Wood Duck Research
Mooresville, N. Carolina
davismc@ieee.org

ABSTRACT

This work is focused on providing emerging technology to respond to the long-recognized need to rapidly and consistently quantify the analyses of a body of psychological tests that are otherwise ostensibly subjective. That need is driven by the increasingly distressing news of the upsurge of self-destructive behaviors: substance abuse, suicide, violent attacks, PTSD symptoms, *etc.* and the need for better early recognition and intervention. The paper reviews one of the important disciplines in psychometrics, projective tests, *e.g*. the Thematic Apperception Test, the Rorschach Test, the Draw-a-Person test. The issues militating against the more widespread use of those instruments are discussed. Three of the most debilitating characteristics of the tests are: high personnel costs of evaluation, lack of consistency in evaluations and imprecise quantification of the results. The paper sets forth emerging technologies that can putatively address all of these issues: Deep Learning, quantum computing, and virtual conversations with interactive computer generated humans. Advances in all of these fields are noted and their applicability to the issues at hand is discussed. The authors present their position that the use of Natural Language Processing techniques, enabled and enhanced by the emerging technologies referenced above now provide a more efficacious way to administer and benefit from projective tests. This would allow neural net training to isolate diagnostic insights. Further, they advance the assertion that such tests will open up entirely new areas of investigation and reveal new areas of insight that have heretofore escaped the attention of mental health professionals. The extensibility of these instruments into other uses is considered and the adaptability into the readers' disciplines is expressed. One of the areas being investigated is the use of these techniques in the recruitment and placement evaluations in police forces. The paper concludes with a short analysis of anticipated progress in each technology.

ABOUT THE AUTHORS

Dan M. Davis is the Director of Technical Initiatives at Wood Duck Research and a consultant for the University of Southern California, focusing on large-scale distributed DoD training, education and conversational mentor avatars. Pre-retirement, for a decade he was the Director of USC’s JESPP project for JFCOM. As the Assistant Director of the Center for Advanced Computing Research at Caltech, he managed Synthetic Forces Express, bringing HPC to DoD simulations. Prior experience includes serving as a Director at the Maui High Performance Computing Center and as a Software Engineer on the ASAS project at the Jet Propulsion Laboratory and Martin Marietta. He has served as the Chairman of the Coalition of Academic Supercomputing Centers and has taught at the undergraduate and graduate levels. He saw duty in Vietnam as a USMC Cryptologist and retired as a Commander, Cryptologic Specialty, U.S.N.R. He received B.A. and J.D. degrees from the University of Colorado in Boulder.

Sanad Z. Rizvi was a visiting Research Assistant at the Institute for Creative Technologies of the University of Southern California. His interests are in Machine Learning Research and its applications on the real world. Given his Computer Science background, he has accumulated good experience in software development through various research activities and programming projects. His expertise in Python has allowed him to make open source contributions to that language. Currently, he is focusing on Machine Learning and Deep Learning research activities. He earned a Bachelor of Technology degree in Computer Science and Engineering from the National Institute of Technology, Mysore, India.

Kenneth Shaw was a visiting Research Assistant at the Institute for Creative Technologies of the University of Southern California. He was selected to participate in the National Science Foundation's Research Experiences for Undergraduates in 2018. His major interests in research include creating life-like interactive experiences using artificial intelligence and motion-planning or decision making for robots. In addition, he is designing and producing new weather balloon electronics and software to better monitor lightning storms at his home institution, as well as consumer-facing utilities on the Blockchain. He is currently studying Computer Engineering and Computer Science in Intelligence at the Georgia Institute of Technology.

Mark C. Davis, Ph.D. is currently retired after careers in the US Navy and as a computer design engineer for both IBM and Lenovo. Rising to the level of Distinguished Engineer at Lenovo, he was responsible for the design of laptop computer cross-disciplinary technology, including PC architecture, embedded systems, open source and virtualization. Previous work was with IBM in the areas of software development and architecture involving security, storage and virtualization. Dr. Davis has been granted well over fifty patents that were filed during his service at both companies. He is a graduate of the Duke University NROTC program and was commissioned as an Ensign, attended nuclear power school, and served as a Submarine Officer for twelve years, including one duty tour as a classroom instructor. He left the service as a Lieutenant Commander to pursue a PhD. Mark holds a BSEE degree from Duke University and a PhD in Computer Science from the University of North Carolina, where his advisor was Professor Fredrick P. Books.

Dan M. Davis
Univ. of Southern California
Long Beach, California
dmdavis@acm.org

Sanad Z. Rizvi & Kenneth Shaw
Univ. of Southern California
Los Angeles, California
mohdsanadzakirizvi@gmail.com
& kshaw@gatech.edu

Mark C. Davis
Wood Duck Research
Mooresville, N. Carolina
davismc@ieee.org

INTRODUCTION

This paper propounds three theses: 1) Projective tests are valuable assets to understanding behavior. 2) Implementation of these tests present significant issues. 3) Merging capabilities in modeling and simulation can both deliver ameliorative solutions for current obstacles and future benefits. The paper begins with a short recounting of why the tests are needed, especially in the population of current and veteran warfighters. The tests themselves are then reviewed and described. There is a section on the obstacles to their widespread and optimal use. The central thrust of the paper is a short list of emerging technologies and techniques that the authors hold will address many of the aforementioned obstacles. Based on experience with both the tests and with the technologies, the paper then turns to potential issues that may arise as the new technologies are implemented in this area. Extensibility and anticipated advances are also analyzed. The paper concludes with a serious look at alternative paths for the future.

BACKGROUND

There is a constant concern with emotional problems in society and in the Department of Defense (Young, 2016). For both prophylactic and therapy purposes, a group of tests have proven effective: the projective tests. These tests have shown utility in assessing a person's ability to avoid psychological disorders, before that person is subjected to stress. They also may be useful for understanding appropriate therapies for those suffering after such exposures. Without such pre-stress filtering, a number of syndromes have become common. Unfortunately, some of these problems are more common amongst DoD warfighters than in the general population, where they are all too common in any case (Kolodny *et al*., 2015). While these problems are manifold, this paper will just focus on a few.

Behavioral Problems

Substance Abuse

While alcohol has long been the most commonly abused substance in both civilian and military environments, various drugs have now overtaken it both in terms of their characteristic of being more debilitating, more addictive, more societally reprehensible, and in terms of their being more likely to be used after the removal of the stress (Baker, 1972). Drug abuse was prevalent in the post-Vietnam era combat veterans (Mintz *et al.*, 1979). While the recovery rate was higher than those addicted in the civilian sector, it did not match the anecdotal dramatic reduction of alcohol abuse by veterans upon their integration back into stateside life. Not only are the drugs debilitating on their own, they are associated with high rates of psychopathology (Volkow, 2001).

Suicide

Suicide is a growing problem in the United States, especially among the young (CDC, 2007) and this trend is also found in DoD personnel (LeardMann e*t al.* 2011). This tragedy is accentuated by the guilt and loss felt within the military community with its long traditions of caring for and supporting unit members. Since 2015, the US Military has lost more than ten times the lives to suicide than it has to combat in the Middle East. The LeardMann article cited above found only a tenuous connection between combat stress and suicide rates, but it was still over twice the civilian population of the same age and gender. They did find a very close correlation between the presence of previous psychopathologies and incidence of suicide.

Violent Attacks and Sexual Assaults

Another area of concern, popularly associated with combat veterans, is that of violent outbursts of veterans leading to injuries or death of multiple victims. One of the authors of this paper used to remind his Introduction to Psychology undergraduate students to always read any correlation data in both directions as a heuristic to ward off unsupportable assumptions as to causality, *e.g.* "Does enlisting in the military lead to violent tendencies or do violent tendencies lead to enlisting in the military?". In either case, a diagnostic device to ascertain such tendencies might be useful in identifying such personal proclivities prior to seeing them act out in destructive ways. While sexual assaults may often be seen as belonging to a different category, there is a widespread assertion that they are more closely related to asexual aggression than would be assumed by the casual observer (Beech *et al*., 2006).

PTSD

Society has recognized for years the destructive nature of combat on military personnel. While the names have evolved over the last century, shell-shock, battle fatigue, and PTSD, many veterans have always had difficulties becoming integrated back into more civil surroundings after their separation from the service. (Owens, *et al*., 2009). Disruptions to life and increased suicide are common effects of PTSD. Its sufferers are often isolated by their own behaviors and are additionally caused to feel shame, self doubt and guilt, a phenomenon that is particularly apparent in males. However, though less likely to experience PTSD-inducing events, females are more likely to develop PTSD than males (Tolin & Foa, 2002).

Projective Tests

A common problem in both ethnographic behavioral studies and in therapeutic diagnostic instruments is that the issues of interest to the test administrator are not explicitly addressed by the subjects or they are intentionally held back for reasons of the subjects' own. One effective way to pierce these vales of silence is the group of tests called projective tests, in that the subjects are asked to project their emotions or opinions onto ambiguous scaffoldings presented by the test administrator. Much used in any number of situations, these tests are well documented in their use to analyze police and other security personnel (Weiss & Inwald, 2018). However, there is a wide range of criticism of these techniques, potentially challenging their validity (Piotrowski, 2015). There are many projective tests, but this paper will suggest three as being well-known examples of the group.

Rorschach Test

In popular culture, the Rorschach, often referred to as "the ink blot test," is the 20th Century systemization of a much older theory. It presents a person an ambiguous image and lets them analyze it. This is seen as informative as to their mental activities. The ink blots were created by Hermann Rorschach, a Jungian Psychoanalyst, who was very artistic, perhaps due to the influence of his father, an art teacher in Switzerland (Searls, 2017.) The subject is shown an inkblot and asked to comment on it. The analyst then evaluates the answer looking at several parameters: location, content, color, motion, originality, and shading. The most often seen example of one is shown in Figure 1. There is much criticism of the test's objectivity, including some court rulings holding that it is not objective (Gacono & Evans, 2008). There are a few alternative evaluation systems and, of course, concomitant discussions about which is more useful (Reese *et al.*, 2014)

Figure 1 - Rorschach Ink Blot Test Image

Thematic Apperception Test (TAT)

The Thematic Apperception Test, commonly referred to as the "TAT," has a similar function and genesis. The creator, Henry Murray, had a graduate student suggest that her observation of her son's using magazine pictures as a stimulus for his fanciful stories might provide some insight into his personality. The concept is that an ambiguous photo, painting or other explicit image (Figure 2,) may spontaneously evoke an expression of the subject's personality that would otherwise not be made apparent to the analyst. It focuses on more dynamic story-telling than the more static "comments" of the Rorschach ink blots (Sandler, 2016). The subject is directed to discuss the background that led up to the picture, what the picture shows is happening at the time, the feelings of people in the picture and what the future holds for the story (Murray, 1943). As with the Rorschach, the TAT is roundly criticized as lacking an objective foundation (Lilienfeld, 2000). One could ask whether the analysis is more reflective of the subject's or the analyst's personalities.

Figure 2 - TAT Sample Image

Draw-a-Person Test

The Draw-a-Person (DAP0 test is an even more active projective test, in that it does not provide the image, but asks the subject to draw a person. Generally attributed to Florence Goodenough and Dale Harris, it has gone through a few iterations of name change to the now gender-neutral Draw-a-Person test. This particular test is used most often for children and adolescents and the basic instructions are minimal, stressing only that the person drawn must be a whole body drawing, *e.g*. feet, legs, arms, body and hands. The scoring method here is more quantified and may contain analysis of up to 71 individual characteristics, as shown in Figure 3. The criticisms appear to be most adamant when these tests were used to measure intelligence, with one critical study finding a fairly low correlation (r = + 0.27) with the standard Wechsler's IQ test (Imuta, *et al*., 2013). On the other hand, a doctoral dissertation from Pakistan purported to find a high correlation between the DAP and school success in non-English-speaking countries (Hasan, 1990). She reports that certain Pakistani cultural characteristics were observed that required additional scoring procedures, which were supplied. One was a focus on clothing accouterments which were more important to Pakistani young women than it would be in European cultures. On the other hand, foci on facial features were not acceptable activities in this Islamic culture. This has led to suggestions that there may be individualized standardizations within distinct groups in order to avoid cross-cultural errors.

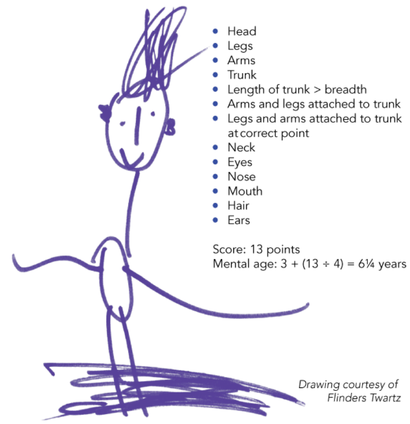

Figure 3 - Draw-a-Person test image (Psych 406, 2015)

Test Implementation Obstacles

High Personnel Costs of Evaluation

The individual cost of any of the above mentioned projective tests is very burdensome. Professionals sufficiently trained and certified to accurately, ethically, and professionally administer and evaluate such tests are rare at best and they are invariably expensive. The fully burdened cost of a therapist, usually either a PhD Psychologist or an MD Psychiatrist, is typically in the couple of hundreds of dollars an hour. Generating the justification for the test will take on the order of an hour, administering the exam will be on the order of another hour and giving a subject the analysis may be yet another hour, so the subject may be looking at something just under $1,000. Psycho-therapy after that will be an additional cost. Those inclined to think "Insurance will cover some of that.", must remember, no matter who writes the check, the consumer pays it, one way or the other, at one time or another. Assessing ~200,000 persons per year, (DoD, 2000) by administering one of these tests to all prospective enlistees to find those less capable of facing combat stresses would be very expensive, it might be on the order $60M every year. Another major cost is the cost in time. Many evaluations not only take time but many tests involve sending the data away to a remote evaluator which may delay either induction or therapy.

Lack of Consistency in Evaluations

As noted above, the process of correlating the results of these tests with any meaningful subject characteristics are in dispute. The agreement between different analysts on the same data is also suspect. Test/retest agreement is sparse. However, the therapists themselves still see these tools as useful. Making the evaluations more rational by depending on more training for evaluators, more formal assessment systems, or some new vision for evaluation systems seems forlorn. Most of the tests described above have been in use for nearly a century, so dramatic improvements in live human evaluations seem unlikely. The postulated bias introduced by the evaluators' individual proclivities would seem to preclude much hope of finding a way to neutralize the natural-born human emotionality of all test evaluators. Any attempt to regularize the evaluative process will also be expensive by virtue of the necessity of both collating and analyzing huge amounts of information. This includes both data on the target behaviors and on the test data. Without such analyses, there would always be the fear that, even with improved reliability and consistency of test evaluations, there might not be any productive impact from the testing. In today's litigious society, there is also the issue of preparing to defend from aggressive civil suits, seeking some legal recourse to real or imagined harms.

Imprecise Quantification of the Results

In all of the behavioral sciences, there is always the problem of quantifying the data. Without quantification, how can one meaningfully understand and convey the insights sought. When one cannot quantify the data, all that is left is emotional intuition, a poor way to make any significant decision. Professor Tetlock of Princeton makes that point in many different ways (Tetlock & Gardner, 2016), but most tellingly, he cites the experience of a CIA officer who prepared a report confidently asserting the occurrence of a truly important event, invasion of Tito's Yugoslavia, as being "a serious probability." When challenged to give the odds of that event by a member of the NSC, he estimated that meant a 65% chance of an invasion. His challengers said that he had thought it meant some other level. So the CIA expert went back to his colleagues and report co-authors and asked them; their quantification ran all the way from 20% to 80%; "serious probability" meant virtually nothing (Kent, 1994). Quantification needs to be a *sine qua non* of any test used to improve the care of DoD personnel or to ensure mission success.

EMERGING TECHNOLOGIES

There are a number of emerging technologies that speak directly to the enhancement of projective test use, including the three discussed above. These technologies seem likely to be useful in overcoming the obstacles that were identified. In different ways, these technologies may also address many analogous issues with other Modeling and Simulation (M&S) methodologies. They may enable and create new methodologies which are currently unforeseen. One of the most exciting new paradigms to be considered is quantum computing and this paper will focus on it and on the ways in which quantum computing can enhance some of the other emerging technologies. Naturally, the use of all of these technologies in M&S may eventually result in extensibility into other areas. It should be remembered that the technologies covered here are all just now coming of age. The stages of the conception, development, maturation and acceptance of new technologies were explored in Professor Clayton Christensen's seminal book *The Innovators' Dilemma* and his theses are adopted in the following sections (Christensen, 1997).

Deep Learning

Deep Learning is a new field of data management built on the old framework of neural nets. Many of the new computationally complex solutions that allow very "human-like" behaviors are made possible by Deep Learning. It has drawn a lot of attention for good reason: it is seen as a potential response to the problems posed by the explosive growth of Big Data (Chen, 2014). It is an emerging technique in machine learning and artificial intelligence. This paper recognizes it as a further extension of earlier research (Demuth, 2014) in the areas of neural networks, evolutionary computing, and data mining. However, in Deep Learning, the refinement is done by several layers of convolution processing, in which some evaluation process sends along only such information that the layer values as have been found to be beneficial. These iterations of processing steps are called hidden layers and are the unique feature of Deep Learning. Deep Learning advocates note that this hierarchical approach to sifting large amounts of data is more in keeping with real word issues, despite its need for additional computational power. Figure 4 is a notional diagram and is intended to show that interconnectivity.

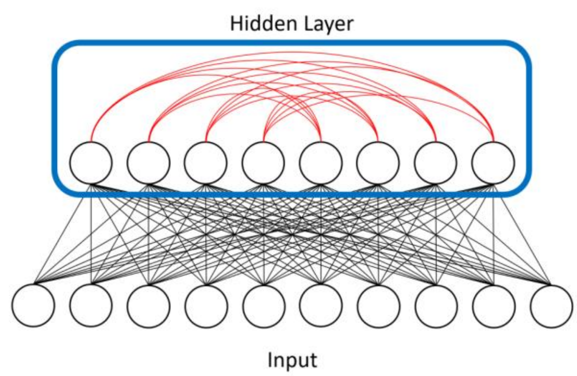

Figure 4 – Notional Diagram of Deep Learning Network (Liu *et al*., 2018)

Deep Learning has been demonstrated as useful in a wide range of problems, including hand writing recognition, object recognition, image classification, speech recognition, understanding natural language text, and adversarial games, (LeCun et al., 2015). A key characteristic of being able to tackle these problems is the need to find complex structures in high-dimensional data. Deep Learning with its multiple layered structure and distributed representation turns out to be well suited for these problems as compared with other machine learning approaches and other manual knowledge engineering techniques.

The current interest in Deep Learning is just the most recent expression of neural network implementations. In the mid-twentieth Century a new type of electronic brain, called the Perceptron, captured the public imagination (Bernstein, 1981). Later, interest in Perceptron fell off when it was pointed out that there were limitations of a single Perceptron, which could only capture simple linearly-separable structures. A few decades later, interest was reawakened as more capable neural network models with multiple layers with multiple neurons per layer were developed. However, this created new problems because multiple layered networks are difficult to train and the computation power was not sufficient at that time. Today, interests in neural network are on the march, as Hinton *et al.* have introduced a novel method to pre-train the network evaluation weights for deep belief networks using Restricted Boltzmann Machines (RBMs) (Hinton, *et al*., 2002). Add to that was the development of improved graphical processing units (GPUs) with Single Instruction Multiple Data (SIMD) capabilities. This reduced the computational bottleneck to a few hours. GPUs are still increasing their power, but also face physical limits to their computational growth (Lucas, 2009).

There are three major Deep Learning approaches (feedforward neural networks, recurrent neural networks, and Boltzmann machines). The Boltzmann machine approach is the most amenable to quantum annealing. Boltzmann machines are probabilistic neural networks that implicitly define probability distribution over the activation states of the neurons in the Deep Learning network. Training a Boltzmann machine requires being able to repeatedly sample from the distribution of activation states. However, sampling for Boltzmann network with loops can be computationally expensive.

The restricted Boltzmann machines can be trained efficiently because they avoid edges within a layer, which enables layer-based approximate sampling of the network. With quantum annealing, there is a potential of efficiently sampling networks with loops. As will be mentioned below, the D-Wave at USC's Quantum Computing Center has an internal, "hard-wired" architecture that is basically a Boltzmann configuration. While this may limit the device from more generalized quantum computing, it is very useful in M&S contexts, as it provides the power needed, in the areas of great interest to the M&S community. This enables the creation of a broader range of networks, with more complex topologies. The current generation of quantum annealers does support complete graphs; they also impose limits on the topology of the intra-layer edges. This is the definition of limited Boltzmann machines (LBM), which are strict supersets of RBMs, and demonstrate the effectiveness of the additional edges. It may be helpful to d
Refer to a diagram showing these basic components of Deep Learning, so Figure 5 is presented here.

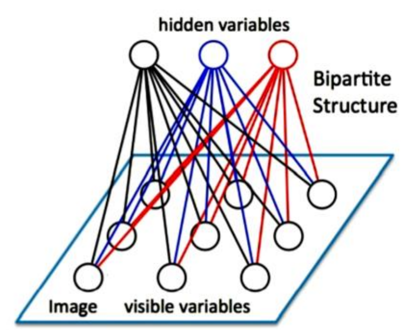

Figure 5 - Notional Image of a Restricted Boltzmann Machine (Pole, 2015)

The over-abundance of data has swamped earlier designs. Deep Learning models require very large datasets to properly train the neural network weights. The activation of a neuron depends on the activation of its neuron neighbors mediated by the weights of the connection. The winner of the 2012 ImageNet Large-Scale Visual Recognition Challenge AlexNet (Krizhevsky & Hinton, 2010) used over 60 million weights with 5 convolutional layers, max pool layers, and 3 fully connected layers. The network was trained on over 15 million images with over 22 thousand labeled categories. The Microsoft ResNet (Jaderberg, *et al.*, 2015) has 152 layers.

Deep Learning does have limitations and there are recent research efforts that are addressing these limitations. One limitation is brittleness. Deep Learning models have demonstrated remarkable ability to perform well in noisy data, for example the ability to recognize objects in photos with busy backgrounds. But, paradoxically it is often quite fragile. Given new classes of objects to recognize, the model must be extensively retrained with many instances of the new objects as well as the previously known objects. In the games area, a Deep Learning-based Alpha Go program can beat top ranked human players, but the number of handicap stones is hardwired. Changing from the default handicap of 7.5 stones requires extensive retraining. An approach to address this limitation is to develop less monolithic deep networks, such as developing visual attention models to incrementally decode images and combining deep networks with artificial reasoning capabilities. Another limitation is the need for large labeled training dataset. Large datasets are often easy to find, but labeling is often a labor intensive process, *cf.* discussion of "sarcasm" recognition below. Researchers are developing unsupervised learning techniques, such as using Boltzmann machine-based auto-encoders; and incorporating artificial intelligence techniques (like reinforcement learning in the AlphGo) to self-generate labeled data. For a more rigorous treatment of these topics, see the work of our colleague Jeremy Liu of USC (Liu *et al*., 2018).

Quantum Computing

As mentioned above, the issues facing the M&S community are challenging enough, but this paper asserts that many missed opportunities to exploit Big Data and new sensors are occurring due to constraints imposed by the digital computing paradigm and the imposition of deterministic methods. The M&S efforts, most particularly those funded by the DoD, are more often directed to gaining insights via analysis and evaluation or provide training by implementing a non-hazardous space in which trainees may practice their skills and try out new approaches. In these cases, *e.g.* battle space simulations, otherwise deterministic solutions are often made more stochastic by introducing random number generators into the parameter values of the algorithms. In one of these battlefield simulation validity discussions, it was noted that if one were simulating the battle of Shiloh, General Grant's order that General Lew Wallace should advance to Pittsburgh Landing would be followed precisely by Wallace, whereas, in fact, the order was not clear Wallace went down the wrong road and was unable to help Grant when needed. As it is in real life, M&S scenarios need to have enough randomness in them to not mislead the analyst, evaluator or trainer that the results obtained mirrored real experiences. With the introduction of randomness, the trainee can get in as many iterations as time is available, but the analysts and evaluators will surly want to know if they have examined the entire space of possibilities. Further, using any neural net training bespeaks the need for a similar range of values to quickly and accurately scan data for meaningful relationships.

In classical digital computing, as the values are carefully constrained (a bit is either a one or a zero), the number of iterations needed to cover the entire compute space one bit change at a time is beyond practicality. Then there comes the concept of quantum computing. First brought into the consciousness of academic researchers by Prof. Richard Feynman of Caltech, the concept of a quantum computer proffered the vision of a computer wherein values were represented by a quantum value, and hence could represent a vast range of values simultaneously (a qubit) and could analyze those with some finite number of other qubits to seek out relationships among them (Feynman, 1985). This operation realization of this was never thought to be an easy task, but in 2007, a Canadian firm, D-Wave was demonstrating a 28 qubit computer and in 2011, one of the first D-Waves to be delivered was stood up at USC with a 128 qubit machine. The USC facility now is running a 2000 qubit plus D-Wave: Figure 6. This machine operates in the neighborhood of fifteen thousandths of a degree above absolute zero and the number of stable qubits varies each time the machine is initiated, as a few of qubits retain enough residual current as to be unusable. To get a quantum machine that is productive at this point in technical development, the D-Wave company opted to have the processor configured as a Quantum Annealer (QA), not a general purpose computer, which in this case is very important to M&S, as that use makes it a very large optimization computer, capable of turning out histograms showing relationships of various parameters. The peaks and valleys of the histogram are areas that would be of great interest to M&S analysts. The current processor design allows examination of a fairly small number (6) of parameters. A Processor under development would increase that number by a factor of two or three; it is anticipated in the next several months.

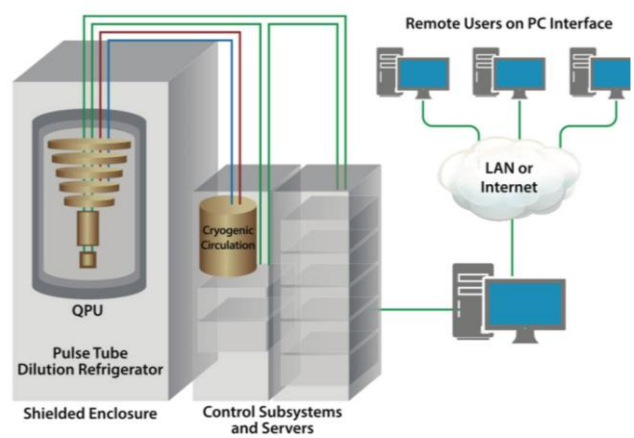

Figure 6 - Block Diagram of QA Installation (D-Wave, 2019)

In 2014, the D-Wave computer showed entanglement, which is taken by most to prove that it has achieved quantum computing, (Lanting, 2014). Again, following the theses of Christensen cited above, it still has not shown the full promise of its miraculous power. The authors of this paper have seen enough hyperbole and dashed, but it is still on a track which is encouraging. Having an operational facility, not a bench-top prototype or a digital simulation of a quantum device, has given the M&S community an opportunity to do work where QA would be advantageous.

Virtual conversations with interactive computer generated humans

There is a MentorPal project at USC which is pursued at the behest of the Office of Naval Research (ONR) to increase the number of students selecting STEM careers. This project is based on earlier USC technology that delivered as conversation-like exchange. To begin, the team recorded a group of about 1,000 short (~1.5 minute) video clips responding to questions about STEM careers, *e.g.* "What do you like about your job?" These clips are then used to present an on-line mentor for high school students via a conversational exchange of data.

Creating a virtual human may seem as simple as remodeling a human using Computer Generated Imagery (CGI.) It takes a great deal of study and effort to implement a Virtual Conversation and the process can consume considerable computing power. The human input and use of technologies that go into the creation of a virtual conversation with lifelike abilities include Natural Language Processing (NLP), machine learning, Virtual Reality (VR), CGI (if animation is involved.), and social stimulation of humans by computer-generated interactions. NLP will be the main focus of this discussion, though the same argument concerning the limits of virtual humans can be made with several of the other components. Natural Language Processing is comprised of the decomposition of language to allow the computer to do useful communications (Chowdhury, 2003). Recent developments in NLP have made significant advances, including breaking down sentences into: parts-of-speech tags, chunks, entity tags, semantic roles, similar words, and the grammatical and semantic elements of a sentence that generate meaning (Collobert & Weston, 2008).

This work is founded on earlier approaches using similar natural language dialog technology; ICT's contribution is called the NPCEditor (Leuski & Traum, 2011), which enables question-answering agents. NPCEditor is one of the two editors used by MentorPal in selecting the optimal video clip in a response to the users' questions. An effective use of the NPCEditor technology was an exhibit at the Boston Museum of Science, The Twins, where museum visitors interacted with dialog-based virtual agents who could answer questions about computer science and about how they worked (Swartout, *et al*. 2010). Later research examined the use of the underlying dialog technology to reproduce the experiences of Holocaust survivors, by producing recorded video clips of the living Holocaust victims to be used as responses, in an attempt to make the conversations more personal than using animated avatars (Traum et al., 2015). These programs relied on a range of interactions and social dynamics. Often they included a live guide who would direct a group of visitors in taking turns asking questions to which the program would respond. In the project here, MentorPal has been designed to focus on lower cost, enhanced program design control, more manageable question sets and improved mentor-like dialog. This "faster-cheaper" approach was intended to enhance adoption of this method by others in similar areas.

This work was all based on previous work and extensive literature on mentoring relying on Artificial Intelligence (AI). The focus of MentorPal is on emulating the experience of an interview with a live mentor, such as the kind a student would have in a counseling office or at a career fair. This can be contrasted with counseling via systems that are designed to help participants working on a specific project, an example of which would be AutoMentor (Wang *et al.*, 2013). MentorPal can also be compared to intelligent mentor agents which gave support to metacognition as part of an open learner model (Dimitrova & Brna, 2016). Among mentoring agents which also handle question-asking and familiarization, *e.g.* the SimCoach system (Rizzo *et al.*, 2011), MentorPal focuses on subjective experiences instead of data. Rather than a general description of STEM careers or fields, it is intended to help learners select a virtual mentor whose experiences resonate with them and then enables them to explore a more realistic vision of a career. A MentorPal in SmartPhone display format is shown in Figure 7.

Figure 7 – iPhone MentorPal display (USC, 2019)

IMPLEMENTATION ISSUES

One of the real challenges in computing has always been the emulation of human behavior with enough cause a rational user to be unable to discern if the interactive agent is a human or a computation device. First popularized by Alan Turning (Saygin *et al*., 2000), but now these concepts are commonly used and the Turing test (Saygin, *et al*., 2000) is "taken" daily by Siri, Alexa, and Google chatbots. Today, were one to try and ferret out if a responder was human, it would be good to probe three areas: recognition of humor, identification of sarcasm, and ability of unseen agent to advance germane and appropriate questions. Here the QA capabilities offer hope, if not promise, in their ability to use techniques like Deep Learning to ferret out that which we as humans can easily detect: humor, sarcasm and facile interactive questions. It is the last issue that is of greatest concern to the M&S professional. There are several issues that would benefit from a computer agent that could, not only answer questions like the chatbots, but could initiate lines of questions based on context and sensor input.

In looking at the stated subject of this paper, the use of virtual humans to administer a projective test would be dependent on a truly conversational agent, not just an animated FAQ utility. There are a number of areas where QA could respond to these needs: analysis of collected data to train the computer to interface more interactively, generation of salient questions, recognition of intent which would otherwise remain opaque to an NLP algorithm, and analysis of all of a test's interactions to establish a conclusory opinion as to the mental state of the subject.

EXTENSIBILITY ANALYSIS

In the future, the community should look forward to the increased use of neural net training to isolate any number of insights. Further, this paper has set forth the position that new technologies may well advance the assertion that such projective tests will open up entirely new areas of investigation and reveal new areas of insight that have heretofore escaped the attention of mental health professionals.

One of the areas recently brought to the USC research team was that of the recruitment and placement evaluations in police force acquisitions. Here, as in defense work, high stress and the criticality of absolute stability in the face of violent disorder and physical threat are highly desired by the public.

ANTICIPATED ADVANCES

The authors all fully anticipate dramatic advances in the areas of Deep Learning, Quantum Annealing and Conversationally Capable Human Computer Agents. There are now new entries into the world of Quantum Computing, including IBM's newest quantum machine with five qubits, one with 16, and one with 20 qubits (Dueck & Pathak, 2018). This entry marks a truly stable beginning of the industry, rather than D-Waves lonely excursion into an otherwise empty discipline. Only time will tell if the area will grow, a la the Christensen hypothesis, or will remain unstable and abandoned.

CONCLUSIONS

The M&S community needs increasing sophistication, ever more ability to identify relationships between behaviors, new ways to handle vast amounts of data, an ability to learn and analyze projective tests, and the capacity of more human-like conversational abilities. The three emerging technologies featured herein hold promise for meeting some known challenges and for creating whole new areas of opportunity to enhance M&S to benefit the defense posture of the U.S. The most speculative, least proven of these is Quantum computing, but it may hold the most promise. Extant hardware and rapidly growing software support should encourage M&S practitioners to seek out new areas where these technologies may serve the needs of their technology consumers. The rapidity of the development evolutions requires periodic re-examination of the costs and benefits of all new technologies. Quantum computing putatively holds promise for advances on all of the technologies addressed in this paper.

ACKNOWLEDGEMENTS

The authors would like to acknowledge that assistance and advice of Jeremy Liu, Ke-Thia Yao and Robert Lucas of USC in helping analyze the progress of quantum computing in the US and Benjamin Nye and William Swartout for their leadership as Principal Investigators of MentorPal.

REFERENCES

Baker, S.L. (1972). US Army Heroin Abuse Identification Program in Vietnam: Implications for a Methadone Program. *American Journal of Public Health*, 62(6), 857-860.

Beech, A. R., Ward, T., & Fisher, D. (2006). The Identification of Sexual and Violent Motivations in Men who Assault Women: Implication for Treatment. *Journal of Interpersonal Violence*, *21*(12), 1635-1653.

Bernstein, Jeremy, (1981), A. I. *New Yorker Magazine,* New York, New York.

Centers for Disease Control and Prevention. (2007). *Suicide Trends among Youths and Young Adults Aged 10-24 years--United States, 1990-2004*. MMWR: Morbidity and Mortality Weekly Report, 56(35), 905-908.

Chen, X. W., & Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, 2, 514.

Chowdhury, G. (2003). Natural Language Processing. *Annual Review of Information Science and Technology*, 37(1), 51-89.

Collobert, R., & Weston, J. (2008). A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In Proceedings of the *25th International Conference on Machine Learning* (pp. 160-167).

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail.* Harvard Business Review Press, Cambridge, Mass.

D-Wave, (2019). *The D-Wave 2000Q, Technology Review*, Retrieved from the internet on 27Jan19 from: https://www.dwavesys.com/sites/default/files/D-Wave%202000Q%20Tech%20Collateral_0117F_0.pdf.

Davis, D.M. (2019). *MentorPal, Virtual Conversations to Enhance STEM Career Selection*, Retrieved on 21 January 2019, from: http://www.hpc-educ.org/MentorPal/Web/MntPal_Index.html.

Demuth, H. B., Beale, M. H., & De Jesús, O. (2014). *Neural Network Design*. Martin Hagan, Ed. Vol. 20. Boston: Pws Pub.

Dimitrova, V., and Brna, P. (2016). From Interactive Open Learner Modeling to Intelligent Mentoring: Style-OLM and Beyond. *International Journal of Artificial Intelligence in Education* 26(1), 332–349.

DoD, (2000). *Department of Defense Overview*, US Department of Defense. Retrieved on 22 January 2019 from: https://comptroller.defense.gov/Portals/45/documents/cfs/fy2000/04_Section_1_FY2000.pdf'

Dueck, G. W., Pathak, A., Rahman, M. M., Shukla, A., & Banerjee, A. (2018). Optimization of Circuits for IBM's five-qubit Quantum Computers. In 2018 *21st Euromicro Conference on Digital System Design (DSD)* (pp. 680-684). IEEE.

Eva Hasan, (1990). *Standardization of Draw-A-Person (DAP) Test in Pakistan*. PhD, Doctoral dissertation, University of Karachi, Pakistan.

Gacono, C. B., & Evans, F. B. (Eds.). (2008). *The Handbook of Forensic Rorschach Assessment*. Taylor & Francis. Milton Park, UK.

Hinton, G. E. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation* 14.8: 1771-1800.

Imuta, K., Scarf, D., Pharo, H., & Hayne, H. (2013). Drawing a Close to the use of Human Figure Drawings as a Projective Measure of Intelligence. *PloS one*, 8(3), e58991.

Jaderberg, M., Simonyan, K. and Zisserman, A. (2015). Spatial Transformer Networks. *Advances in Neural Information Processing Systems*. (pp. 2017-2025).

Kent, S. (1994). Words of Estimative Probability, in Sherman Kent and the *Board of National Estimates*. Ed. Donald P. Steury, pp. 134–35.

Kolodny, A., Courtwright, D. T., Hwang, C. S., Kreiner, P., Eadie, J. L., Clark, T. W., & Alexander, G. C. (2015). The Prescription Opioid and Heroin Crisis: a Public Health Approach to an Epidemic of Addiction. *Annual Review of Public Health*, 36, 559-574.

Krizhevsky, A. & Hinton, G.E. (2010). Factored 3-way Restricted Boltzmann Machines for Modeling Natural Images. *International Conference on Artificial Intelligence and Statistics.* Las Vegas, Nevada.

Lanting, T., Przybysz, A. J., Smirnov, A. Y., Spedalieri, F. M., Amin, M. H., Berkley, A. J., ... & Dickson, N. (2014). Entanglement in a Quantum Annealing Processor. *Physical Review X*, 4(2), 021041.

LeardMann, C. A., Powell, T. M., Smith, T. C., Bell, M. R., Smith, B., Boyko, E. J., ... & Hoge, C. W. (2013). Risk Factors Associated with Suicide in Current and Former US Military Personnel. *JAMA*, *310*(5), 496.

LeCun, Y., Bengio, Y. & Hinton G.E. (2015). Deep Learning. *Nature* 521.7553: 436-444.

Leuski, A., and Traum, D. (2011). NPCEditor: Creating Virtual Human Dialogue using Information Retrieval Techniques. *AI Magazine* 32(2), 42–56.

Liu, J., Spedalieri, F. M., Yao, K. T., Potok, T. E., Schuman, C., Young, S., ... & Chamka, G. (2018). Adiabatic Quantum Computation Applied to Deep Learning Networks. *Entropy*, 20(5), 380.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The Scientific Status of Projective Techniques. *Psychological Science in the Public Interest*, 1(2), 27-66.

Lucas, R. F., Davis, D. M., Wagenbreth, G., & Tran, J. J., (2009), Operational Use of a Large GPGPU-Enhanced Linux Cluster. in the Proceedings of the *High Performance Computing Modernization Program Users' Group Conference*, San Diego, California, 2009.

Mintz, J., O'Brien, C. P., & Pomerantz, B. (1979). The Impact of Vietnam Service on Heroin-addicted Veterans. *The American Journal of Drug and Alcohol Abuse*, 6(1), 39-52.

Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, MA, US: Harvard University Press.

Owens, G. P., Steger, M. F., Whitesell, A. A., & Herrera, C. J. (2009). Post-Traumatic Stress Disorder, Guilt, Depression, and Meaning in Life among Military Veterans. *Journal of Traumatic Stress: Official Publication of the International Society for Traumatic Stress Studies*, *22*(6), 654-657.

Piotrowski, C. (2015). Projective Techniques Usage Worldwide: A Review of Applied Settings 1995-2015. *Journal of the Indian Academy of Applied Psychology*, *41*(3), 9-19.

Pole, I., (2015), Restricted Boltzmann Machine - A comprehensive study with a focus on Deep Belief Networks. A Paper submitted to an *Advanced Seminar in Computational Intelligence*, *Frankfurt University of Applied Sciences*, Frankfurt, Germany.

Psych 406, (2015). *Let's Draw Out our Intelligence*, Last Remnant blog, Retrieved from the Internet on 23 January 2019 from: https://medium.com/psyc-406-2015/let-s-draw-out-our-intelligence-3137b9b59395.

Reese, J. B., Viglione, D. J., & Giromini, L. (2014). A Comparison between Comprehensive System and an Early Version of the Rorschach Performance Assessment System Administration with Outpatient Children and Adolescents. *Journal of Personality Assessment*, 96(5), 515-522.

Rizzo, A., Lange, B., Buckwalter, J.G., Forbell, E., Kim, J., Sagae, K., Williams, J., Difede, J., Rothbaum, B.O., Reger, G., (2016). Simcoach: an Intelligent Virtual Human System for Providing Healthcare Information and Support. *International Journal on Disability and Human Development* 10(4), 277–281.

Saygin, A. P., Cicekli, I., & Akman, V. (2000). Turing test: 50 years later. *Minds and Machines*, *10*(4), 463.

Searls, D. (2017). *The Inkblots: Hermann Rorschach, his Iconic Test, and the Power of Seeing*. Broadway Books. New York, NY.

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., ... & Lane, C. (2010). Ada and Grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents* (pp. 286-300). Springer, Berlin, Heidelberg.

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The Art and Science of Prediction*. Random House. New York, NY,

Tolin, D. F., & Foa, E. B. (2002). Gender and PTSD: A Cognitive Model. In R. Kimerling, P. Ouimette, & J. Wolfe (Eds.), *Gender and PTSD* (pp. 76-97). New York, NY, US: Guilford Press.

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., ... Haase, K., 2015: New Dimensions in Testimony: Digitally Preserving a Holocaust Survivors Interactive Storytelling. In: *International Conference on Interactive Digital Storytelling*. pp. 269–281. Springer.

Volkow, N. D. (2001). Drug Abuse and Mental Illness: Progress in Understanding Comorbidity. *American Journal of Psychiatry*, 158(8), 1181-1183.

Weiss, P. A., & Inwald, R. (2018). A Brief History of Personality Assessment in Police Psychology: 1916–2008. *Journal of Police and Criminal Psychology*, 1-12.

Young, J.L. (2016). Military Mental Health, Understanding the Crisis, *Psychology Today*, 10 June 2016, United States Blog, Retrieved on 15 January 2019 from: https://www.psychologytoday.com/us/blog/when-your-adult-child-breaks-your-heart/201606/military-mental-health.

Wang, J., Li, H., Cai, Z., Keshtkar, F., Graesser, A., Shaffer, D.W. 2013. Automentor: Artificial Intelligent Mentor in Educational Game. In: *International Conference on Artificial Intelligence in Education*. pp. 940–941. Springer.

Rizzo, A., Lange, B., Buckwalter, J.G., Forbell, E., Kim, J., Sagae, K., Williams, J., Difede, J., Rothbaum, B.O., Reger, G., (2011). Simcoach: an Intelligent Virtual Human System for Providing Healthcare Information and Support. *International Journal on Disability and Human Development* 10(4), 277–281.

USC, (2019). *MentorPal, SmartPhone Format*. Information Sciences Institute, USC, Playa Vista California.