BrainCallus
Gaming
Project

Using Games to Improve Psychiatric Decision-Making

Project Proposal

Erik J. Schlicht, PhD

BrainCallus Gaming Project

October 22, 2020

Executive Summary

Benefit Statement

This proposal provides a method for quantifying symptoms of mental health disorders through the use of computational games and machine learning. The proposed method has the following benefits:

1. **Continuous Measurements.** By leveraging mobile games, it allows the patient to participate in measurement activities outside the clinic. This provides the physician greater insight into situations that contribute to increased symptoms.
2. **Implicit (Nonverbal) Measures.** Many times patients are unwilling or unable to report when they are experiencing symptoms related to their disorder. Computational games allow for *implicit* measurements of patient symptoms to be obtained, providing the physician with otherwise unavailable metrics that are useful to inform diagnostic or treatment decisions.
3. **Compelling Data Elicitation.** The use of games provides an approachable and enjoyable measurement activity. This allows for *targeted data* to be collected that are easily interpreted. Moreover, these data are useful for training machine learning algorithms that predict clinically relevant information, and promote Just-In-Time Adaptive Interventions (JITAIs).

1 Significance

Cognitive deficits are a hallmark of many psychiatric disorders [**1**] but they are difficult to evaluate outside the clinic. This is because evaluating a patient's cognitive state is currently done by the physician, leveraging psychometric inventories that are expensive in both time and money.

The targeted data obtained through traditional psychometric inventories provide temporal snapshots of the patient's symptoms, offering little insight into how symptoms dynamically change across patient context (e.g., time and location). Moreover, these data require the patient to consciously report symptoms, which can be problematic in cases when the patient is unable or unwilling to self-report. Finally, these measurement activities are not enjoyable, limiting the willingness of patients to participate in repeated measurements.

These shortcomings result in challenges when diagnosing and treating mental health disorders. For example, the lack of objective data provided through self-report can lead to misdiagnosis because the patient is unable to accurately describe symptoms they were experiencing between clinical visits. Moreover, the effectiveness of treatments are difficult to evaluate since there are few objective metrics that can be used to compare baseline and treatment conditions. Therefore, developing methods that are capable of producing frequent objective measurements of patient symptoms would vastly improve the ability of physicians to diagnose and treat mental health disorders.

2 Innovation

Contemporary methods for detecting mental health disorders leverage *passive data* (e.g., text and voice) from devices that are used throughout the patient's daily routine (Figure 1). These techniques overcome the limitations associated with self-report methods (Table 1), making it a promising line of investigation. Moreover, the large amount of data produced by these devices provide a mechanism for training machine learning algorithms that predict the presence or decline of a mental health disorder. Presumably, this offers physicians an opportunity to administer treatment with greater responsiveness.

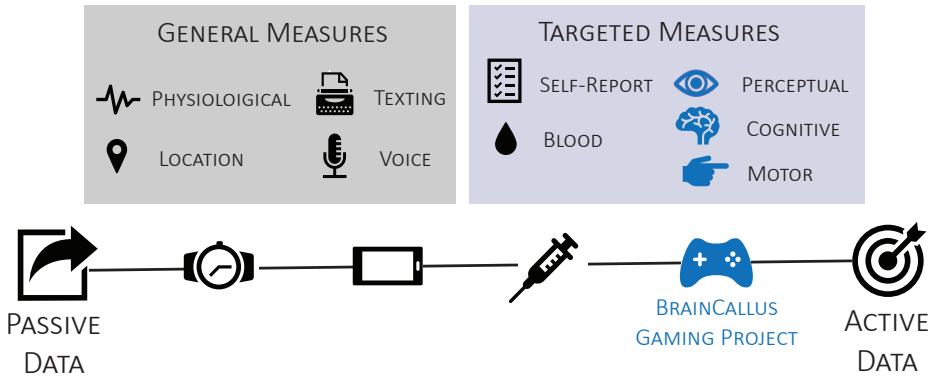


Figure 1: Spectrum of data used in mental health treatment

However, algorithms trained using passive data only detect the physical byproducts of mental health disorders, offering little insight into the symptoms that the patient is experiencing (e.g., decreased attention or increased impulsivity). As a result, it is difficult for physicians to interpret the algorithm's output and act on it with effective treatment. Moreover, the presence of confounding factors (e.g., demographic) in this line of research has been problematic, limiting the potential for these technologies [**2**].

As a result of the above limitations, the goal of the BrainCallus Gaming Project is to develop a method that is capable of *implicitly capturing a large amount of targeted data*. This would allow for machine learning

Data Type	Actionable	Ample Data	Implicit	Enjoyable
Passive		✓	✓	✓
Self-Report	✓			
Blood	✓		✓	
BrainCallus	✓	✓	✓	✓

Table 1: Table depicting the benefits and limitations of current approaches

algorithms to be developed that predict clinically relevant information, while leveraging data that provides actionable insight (Table 1).

Accomplishing this goal requires a system architecture where game metrics are converted to clinically actionable information through the use of machine learning algorithms, while allowing physician feedback to inform gaming modules that the patient will participate in between clinical visits (Figure 2). Notice this architecture provides a rapid feedback-loop between physicians and patients, enabling responsive and effective therapy.

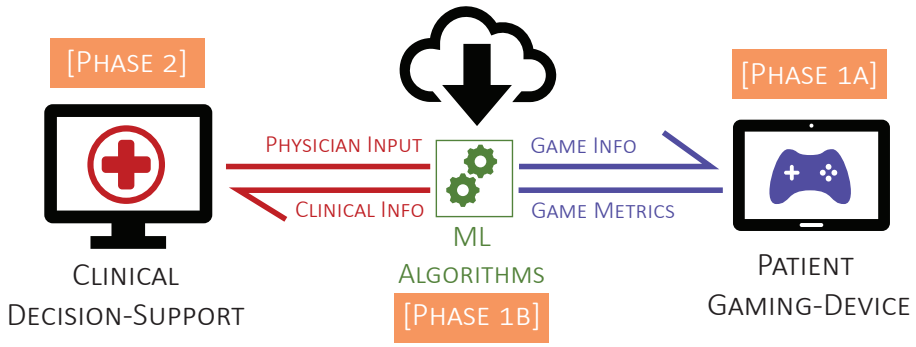


Figure 2: High-level architecture of the system to be prototyped in this proposal

This proposal aims to develop a prototype of the desired architecture through each of the phases of the project. To that end, the next section describes a high-level overview of the computational games proposed in this effort, focusing on how the game metrics correspond to common symptoms of psychiatric disorders. The remaining sections offer insight into how the metrics produced by games can be utilized by machine learning algorithms to provide clinically relevant and actionable information.

3 Phase Ia: Computational Games for Quantifying Symptoms

The computational games that are produced in this effort will provide metrics that quantify common symptoms of psychiatric disorders (Table 2). Developing computational games that are useful for computing metrics of patient symptoms requires incorporating techniques from computational cognitive science, behavioral economics, and computational sensorimotor control. Moreover, producing algorithms that use these metrics as features to produce clinically relevant information requires integrating techniques from machine learning and Artificial Intelligence (AI). This section will describe these processes in greater detail.

Computational games are designed to produce data that is useful for quantifying player performance. More specifically, they contain gaming modules that are created to quantify each component of the human perception-action cycle [**3**]. Perception-action cycles are robust representations that are used to understand human performance across disparate operational settings. Intuitively, perception-action cycles reflect that humans *perceive* the state of the world and use this information to make a *decision* about the best

Symptom	Game Module	Game Metric
Difficulty Concentrating	Chameleon Corral	Attention
Negative Outlook	Face Feline	Bias
Increased Impulsivity	Cash Cow	Risk Aversion
Sensitivity to Loss	Cash Cow	Loss Aversion
Psychomotor Slowing	Duck Dash	Reaction Time

Table 2: Common symptoms of psychiatric disorders and corresponding game metrics

course of action to accomplish their goals. Then, they *act* on the world based on their decision, which changes the state of the world and the cycle is repeated.

Figure 3: Brain Barn contains a series of prototype computational games used in this proposal

Since perception-action cycles can encapsulate human performance across several situations, developing metrics that quantify patient symptoms would involve producing metrics that quantify each level of the perception-action cycle. Therefore, the *BrainCallus Gaming Project* seeks to develop computational games that have a compelling overall storyline (i.e., *Brain Barn*), while containing modules that isolate and quantify each component of the perception-action cycle. This allows the game to produce data useful for quantifying patient symptoms, while still being enjoyable to play. Moreover, **this initial sampling of computational games will provide proof-of-concept that games are capable of producing metrics with clinical significance**. The next few sections will highlight performance metrics that can be obtained within each level of the perception-action cycle.

3.1 Perceptual Metrics

Perceptual deficits are prevalent in several psychiatric disorders, such as schizophrenia [**4**]. In order to quantify these deficits within the game, modules will be developed that are inspired from computational experiments in perceptual science [**5**]. This will allow for metrics to be produced that reflect high-level perceptual performance, such as perceptual [**5**] and emotional [**6**] inference, in addition to low-level metrics that evaluate attention [**7**] and perceptual memory [**8**].

3.1.1 Low-Level Attention

As part of the Brain Barn game series, we have already developed (using Unreal Engine [**9**]) a module that is capable of quantifying player attention. The Chameleon Corral game is based on psychophysical research using multi-object tracking [**10**] and is capable of providing metrics of player attention by participating in gaming modules that last approximately 5 minutes. Difficulty in the game is modified by increasing the

Figure 4: Chameleon Corral is capable of quantifying player attention

number of targets and distractors, in addition to changing the color of the chameleons to a color similar to the background. The game also has meters that track progress, performance and status. The meters incentivise players to try hard during play, in addition to returning to play the game repeatedly.

Prediction: Since attention can be adversely impacted by mental health disorders, we predict that those with disorders will have lower attention scores, particularly on levels with high difficulty.

3.1.2 High-Level Emotional Inference

The Brain Barn series will also develop a perceptual game (i.e., Face Feline) that is capable of quantifying high-level emotional inference. It is known that individuals with mental health disorders have a bias for disproportionately processing negative facial emotions [**11**], so this gaming component will produce metrics that quantify this bias.

The mechanics of this game will be similar a match-to-sample task [**12**], using faces that have been normed across subjective ratings of emotional content [**6**]. Participants will be presented a neutral exemplar face and asked to select the face (from three options) that best matches the emotional content of the exemplar face. These experimental mechanics will be reproduced in a gaming module, ensuring that participation is enjoyable and lasts fewer than 5 minutes.

Prediction: Since mental health conditions can lead to a negative emotional bias, we predict that participants with a disorder will select faces with negative emotional content at greater probability than those without, when attempting to match to an emotionally neutral face.

3.2 Decision Metrics

The decision-making modules are motivated by experiments in behavioral economics [**13**]. As a result, metrics such as risk and loss-aversion can be computed, in addition to metrics that reflect the ability of the player to delay reward [**14**]. Together, these metrics will reflect the player's impulsivity, which is known to be associated with several psychiatric disorders [**15**].

3.2.1 Risk- and Loss-Aversion Estimates

As part of the Brain Barn series, we have already developed a game capable of quantifying player risk- and loss-aversion tendencies. The Cash Cow game requires players to choose between a stochastic gamble (in the Barn A casino), or a sure outcome (in the Barn B casino). These dynamics are identical to classic paradigms in behavioral economics thereby allowing us to use a softmax utility model to quantify player risk- and loss-aversion [**16**]. Similar to Chameleon Corral, the Cash Cow game also has progress, money and status meters to incentivise players.

Figure 5: Cash Cow game is capable of quantifying player decisions

Prediction: Since mental health disorders can lead to impulsivity, we predict that those with disorders will display risk-seeking behavior, reflected in their risk-aversion parameters. Moreover, we predict that those with disorders will also exhibit increased sensitivity to losses, which will be apparent in their loss-aversion parameter.

3.3 Sensorimotor Metrics

Psychogenic movement disorders and psychomotor slowing are a hallmarks of several psychiatric disorders [**17**, **18**]. The Duck Dash gaming module in the Brain Barn series will quantify player movements. More specifically, the module will produce metrics that evaluate changes in motor variability [**18**] and optimality of action selection [**19**] across time, providing insight into motor performance. This will be the final installment of the Brain Barn Game series.

Prediction: Since mental health disorders can lead to psychomotor slowing [**18**], we predict that those with disorders will exhibit slower reaction times than those without disorders. Moreover, we will be able to explore if disorders lead to differences in movement variability or optimality in target selection. We currently do not have strong predictions regarding the direction of differences (if any) for these metrics, but they are still worth exploring since the game will be able to capture these measures.

This section provided an overview of how the Brain Barn gaming modules will produce metrics that quantify common symptoms of psychiatric disorders (Table 2). The next section will highlight how these game metrics can be used as features for training machine learning algorithms that predict clinically relevant information.

4 Phase Ib: Machine Learning Algorithm Development

The second component to prototyping the system depicted in Figure 2 is to evaluate the ability of machine learning algorithms to convert gaming metrics into clinically relevant information. This section will describe how this goal will be accomplished through the scope of this proposal.

4.1 Obtaining a Large Player Population

A necessary first step is to developing our system is to identify a large population of players that are capable of accomplishing two fundamental tasks:

- Beta-test games produced in Section 3 to assure they are working as intended and are enjoyable to play
- Provide labels (e.g., psychiatric disorder present/absent) to train the machine learning classifiers that will be evaluated in this section

Fortunately, the Beta-Family service [**20**] has approximately two-hundred and fifty thousand players available to test the Brain Barn game. Moreover, these players can be selected based on demographic information, in addition to other relevant factors that can be screened based on pre-game survey results (e.g., currently being treated for a mental health disorder). Taken together, this suggests that the Beta-Family site will provide all the necessary functionality for accomplishing the goals of this proposal, and will be used for this effort.

4.2 Interpreting Symptom Scores

Since game metrics can be interpreted with respect to patient symptoms (Table 2), it is desirable to provide the physician with explicit information about how to interpret each score (Figure 6). To that end, game metric norms will be developed that allow percentile scores to be calculated, enabling direct interpretation of symptom risk.

This provides the physician with symptom understanding at-a-glance and could assist in clinical decision-making alone by providing an easy to interpret symptom information. Moreover, it allows a mechanism for signaling a Just-In-Time Adaptive Intervention (e.g., high risk symptom percentile threshold) that could offer a recommended gaming module assignment to gain additional information about the area of concern (Figure 6).

4.3 Evaluating Predictive Models

In addition to providing the physician with actionable information about patient symptoms, there is great potential for training machine learning algorithms that predict clinically relevant information by leveraging game metrics as features. This section will briefly highlight efforts that will be made to explore this potential.

First, the predictive performance of different algorithms will be evaluated by their ability to predict the probability of a mental health disorder, using the game metrics as features. This will accomplished by using the binary labels provided from beta-testing (condition present/not-present) to train our machine learning algorithms. Since this is a supervised binary classification problem, appropriate machine learning models will be trained (e.g., Logistic, Bayesian, SVM, Random Forest) and evaluated (i.e., using AUC). Moreover, Deep Learning methods will be trained and their predictive performance compared to more traditional methods, mentioned above.

Furthermore, each patient's time series data will be used to evaluate the ability of algorithms (e.g., LSTM RNNs) to predict symptom changes over time. If successful, this will allow physicians insight into when patient symptoms are likely to flare-up, allowing them the opportunity to prepare the patient and also try to assess the root-cause (in combination with location data) for the increased symptoms.

Finally, we will investigate the ability of different sampling algorithms (e.g., cross-entropy methods) to provide maximum-information recommendations for gaming modules that reduce uncertainty about high-risk symptoms. If successful, these algorithms can be used to provide recommendations to the physician regarding gaming modules that offer the most information about areas of patient concern (Figure 6). This would be an approximation of a Just-In-Time Adaptive Intervention (JITAI).

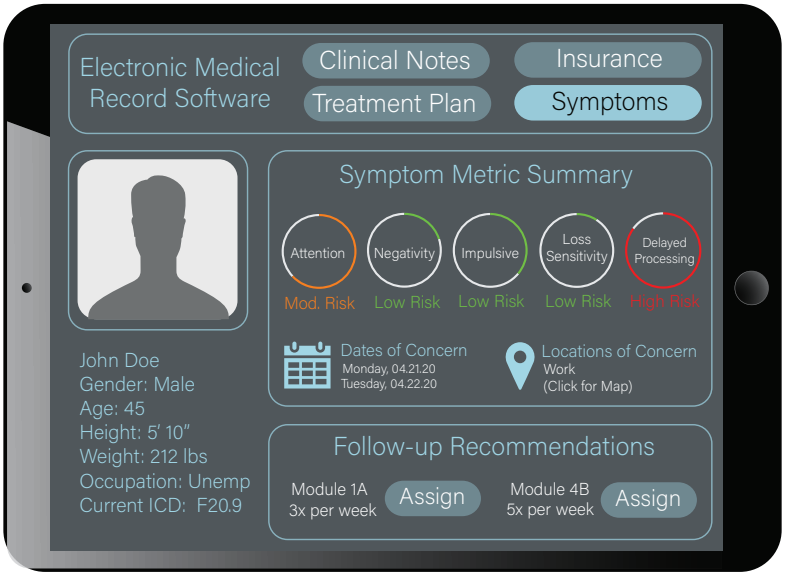
Overall, the goal of the work in this section is to perform a preliminary evaluation of the potential for game metrics to provide clinically relevant information. The next phase of this project will investigate this potential with a greater degree of precision by directly incorporating physician input.

5 Phase II: Developing Clinical Decision-Support

The goal of decision-support systems is to provide physicians with information that is relevant to their clinical context (Figure 6). In a quantitative sense, these systems should provide observations that maximize

the physician's ability to infer hidden clinical states, in addition to offering recommendations that are most efficient at obtaining clinical goals [**3**].

Figure 6: Prototype of a clinical decision-support system using the symptom metrics from Phase I



During Phase II efforts, rigorous steps will be taken to develop effective decision-support technology. First, we will explicitly define each clinical context by using physician input. For every context, we will delineate the hidden clinical states that must be inferred. Then, we will determine all of the observations allowed by the available data to infer each hidden state. Finally, from all of the potential observations, we will provide only the minimum set that provides maximum reliability through our decision-support system.

The information provided through the decision-support system will be visualized in a manner that enables rapid fused beliefs to be formed about the current clinical state. Once a belief has been formed, the system will prescribe the clinical action that allows the most efficient path from the current state to the desired state.

Although the details of this process are beyond the scope of this Phase I proposal, collaborations with clinical partners will be secured before starting Phase II efforts. More specifically, we will present our Phase I findings to suitable partners and pursue the opportunity that maximizes our ability to effectively conduct these steps.

Overall, the BrainCallus Gaming Project provides a feasible and innovative solution to overcoming limitations inherent to current methods. It uses computational games to implicitly produce targeted metrics that quantify patient symptoms outside the clinic. Then, it leverages machine learning algorithms to provide clinically relevant information that is displayed to the physician using decision-support software. We are excited to explore the potential of this technology to improve psychiatric decision-making.

‘

6 Proposal Author

Erik J. Schlicht, PhD. (PI and PM)
Dr. Schlicht is the founder of the BrainCallus Gaming Project. He conducted research at Harvard University, MIT, Caltech and the University of Minnesota, where he used machine learning and AI to quantify human performance under uncertainty and risk. He received his PhD from the University in Minnesota in Cognitive and Brain Sciences, with a minor in Human Factors.

At MIT Lincoln Laboratory, he was a core member of a team who developed methods utilizing Serious Games to quantify operational decision-making under risk [**21**]. While a postdocoral researcher between Harvard University and Caltech, Dr. Schlicht created a simplified poker task to quantify decision-making in a competitive (zero-sum) task. This effort resulted in a publication [**16**] that ranks in the top 5% of all research output, according to metrics by Altmetric [**22**].

References

[1] Trivedi, J.K. (2006). Cognitive deficits in psychiatric disorders: Current status. *Indian Journal of Psychiatry*, 48, pp.10-20

[2] Wei, p. & et al. (2019). Re-examining the robustness of voice features in predicting depression: Compared with baseline of confounders. *Games for Health Journal*, 6

[3] Schlicht, E.J.(2018). Designing effective decision-support. *The Machine Learning Conference*. https://mlconf.com/guest-blog-by-erik-schlicht-founder-of-the-computational-cognition-group-llc/

[4] Uhlhaas, P.J. & Mishara, A. (2006). Perceptual Anomalies in Schizophrenia: Integrating Phenomenology and Cognitive Neuroscience. *Schizophrenia Bulletin*, 33, pp. 142-156.

[5] Knill, D. (2008). Perception as Bayesian Inference. *Cambridge University Press*, ASIN B008GZZHC2.

[6] Said, C.P., Haxby, J.V. & Todorov, A. (2017). Brain systems for assessing the affective value of faces. *Philos Trans R Soc Lond B Biol Sci*, 366, pp. 1660–1670.

[7] Itti, L. & Borji, A. (2014). Computational models of attention. *Cognitive Neuroscience: The Biology of the Mind*, 5, pp. 1-10.

[8] XiaoLan, F., et al. (2014). A computational cognition model of perception, memory, and judgment. *Science China Information Sciences*, 57, pp. 1-15.

[9] Unreal Engine. *The most powerful real-time 3D creation platform*. https://www.unrealengine.com/en-US/

[10] Meyerhodd, H.S., Papenmeier, F. & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, doi:10.3390/ijerph17010330

[11] Simcock, G., et al. (2020). Associations between Facial Emotion Recognition and Mental Health in Early Adolescence. *Int J Environ Res Public Health*,17, pp. 1255–1274.

[12] Bobak, A.K., Hancock, P.J.B. & Bate, S. (2015). Super-recognisers in Action: Evidence from Face-matching and Face Memory Tasks. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3170

[13] Camerer, C. (2011). Behavioral Game Theory: Experiments in Strategic Interaction. *Princeton University Press*, ASIN B005KLQXNC.

[14] Story, G.W., Moutoussis, M., & Dolan, R.J. (2015). A Computational Analysis of Aberrant Delay Discounting in Psychiatric Disorders. *Frontiers in Psychology*, 6, p. 1948.

[15] Kisa C., Yildirim S.G., & Göka E. (2005). Impulsivity and mental disorders. *Turk Psikiyatri Derg*, 16, pp. 46-54.

[16] Schlicht, E.J., et al. (2010). Human Wagering Behavior Depends on Opponents' Faces. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0011663

[17] Cummins G., Zandi M., & Barker R.A. (2015). Movement disorders and psychiatry. *Neurol Clin Pract.*, 5, pp. 143-149.

[18] Djamila, B., et al. (2013). Psychomotor Retardation in Depression: A Systematic Review of Diagnostic, Pathophysiologic, and Therapeutic Implications, *Biomed Res Int*, doi:10.1155/2013/158746.

[19] Schlicht E.J. (2007). Statistical decision theory for human perception-action cycles. *University of Minnesota Doctoral Thesis*.

[20] Beta-Family. *Find and engage app testers*. https://betafamily.com/

[21] Alhadeff, E.(2013). Serious Games at MIT Lincoln Laboratory. *Serious Game Market*. https://www.seriousgamemarket.com/2013/03/serious-games-at-mit-lincoln-lab.html

[22] Altmetric Rating for Human Wagering Behavior Depends on Opponents' Faces. *Altmetric Webpage*. https://www.altmetric.com/details/180954