The Yale Undergraduate Research Journal

Volume 2
Issue 1 *Spring 2021*

Article 35

2021

Examining the Viability of Computational Psychiatry: Approaches into the Future

Mitchell Ostrow
Yale University

Examining the Viability of Computational Psychiatry: Approaches into the Future

Cover Page Footnote

The author thanks Professor Avram Holmes for rigorous discussion and guidance both during the development of this project and afterwards.

Examining the Viability of Computational Psychiatry: Approaches into the Future

By Mitchell Ostrow[1]

[1]*Program in Neuroscience, Yale University; Department of Statistics and Data Science, Yale University*

ABSTRACT

As modern medicine becomes increasingly personalized, psychiatry lags behind, using poorly-understood drugs and therapies to treat mental disorders. With the advent of methods that capture large quantities of data, such as genome-wide analyses or fMRI, machine learning (ML) approaches have become prominent in neuroscience. This is promising for studying the brain's function, but perhaps more importantly, these techniques can potentially predict the onset of disorder and treatment response. Experimental approaches that use naive machine learning algorithms have dominated research in computational psychiatry over the past decade. In a critical review and analysis, I argue that biologically realistic approaches will be more effective in clinical practice, and research trends should reflect this. Hybrid models are considered, and a brief case study on major depressive disorder is presented. Finally, I propose a novel four-step approach for the future implementation of computational methods in psychiatric clinics.

INTRODUCTION

Psychiatrists traditionally utilize behavior and psychology in the clinic but have long sought to ground the practice in biology. Unfortunately, contemporary research has yet to translate, due to the inevitable truth that the brain cannot be carved at its joints. This means that enormous neural complexity has prevented modern methods from sufficiently elucidating pathophysiological processes. Similarly, the conceptual bridges between cellular biology, systems neuroscience, and behavior are shaky given the limits of neuroscientific theory as well as data collection capabilities.

There exist only a handful of mechanistic theories of dysfunction in mental illness, such as the dopamine or glutamate hypothesis in schizophrenia (Seeman, 1987; Gordon, 2010). These first steps have refuted the contemporary understanding disorders as having one-to-one biological mappings. The practical conception of disorder is defined in the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-V; American Psychiatric Association, 2013) on the basis of symptom presentation. This is likely incorrect for a few reasons. First, a DSM-V disorder might encapsulate subtypes with varying etiologies. Second, the prevalence of comorbidities such as depression and anxiety (according to Brady et al., 1992, almost 62%) likely indicates overlapping or interacting neural correlates of various pathologies. Pervasive comorbidity and heterogeneity in psychologically-defined disorders have provoked the Research Domain Criteria (RDoC) approach, which seeks to rebuild psychiatric disorders from biology upwards (Insel et al., 2010).

Neuroscience faces an accelerating deluge of information which analytic trends have not reflected. Imaging studies often fail to replicate (Jahanshad, 2019). A common concern has been limited sample size; open-source initiatives seek to mitigate this issue by sharing data (Poldrack & Gorgolewski, 2019). Data preprocessing techniques in fMRI vary between sites, which have significant influence on results (Smith & Nichols, 2018). Additionally, statistical techniques used in much of neuroscience are simply out of date. Linear models are interpretable, but neural systems are highly nonlinear and analyses should reflect this (Friston, 2004).

In this paper, I analyze the effectiveness of computational tools for clinical psychiatry research and practice. Although the clinical implementation of these methods is the ultimate goal, I seek to examine the viability of computational psychiatry as a research method for developing precision psychiatry. This is not to say that we cannot examine how they might fit into the clinic itself. The beauty of these tools is that they can often be used predictively and to generate understanding, allowing for usage in both research and practice. However, the field is a long way from success in either domain. Thus, I seek to provide a comprehensive review of computational methods applied to psychiatry in general in the hopes of providing a clearer picture about where the field is headed.

BRIEFLY: WHAT IS COMPUTATIONAL PSYCHIATRY?

Computational psychiatry as a discipline is divided into theory- and data-driven approaches (Huys et al., 2016). Theory-driven approaches model the biological processes that generate dysfunction, whereas data-driven models remain agnostic to underlying causes, utilizing statistical trends in data to make inferences on new samples (Bennet et al., 2019). Montague et al. (2012) define a framework for five subdomains: a) data-mining, modeling and phenotyping, b) producing new biological hypotheses, c) large-scale data

Ostrow | Computational Psychiatry

sharing, d) biomarker discovery, and e) application to therapeutics. Data-driven models often utilize supervised ML methods, which have the goal of predicting labels from labelled data (Shatte et al., 2019). Another type, unsupervised learning, extracts statistical patterns from data. Principal component analysis, a linear technique that reduces the dimensions of data into sub-components, is popular due to its interpretability (Drysdale et al., 2017; Bondar et al., 2020). In theoretical work, researchers seek to infer cognitive or neurological states from behavior or neuroimaging. Models, such as dynamic causal models (see Theoretical Approaches) seek to represent disorder of biophysical processes and are often utilized in conjunction with neuroimaging (Friston et al., 2017).

The bridge across the chasm between computational psychiatry research and the clinic is neither stable nor complete. In this analogy, theory-driven approaches might be the slow yet precise construction of an overpass, whereas data-driven approaches are a ramp to facilitate a motorcycle jump, Evil Knievel-style. The latter method is faster, but risk and uncertainty are significantly higher as they do not refer to any underlying disorder. Additionally, data-driven approaches are plagued with poor methodology (Rutledge et al., 2019).

Computational approaches can be quite relevant to some psychiatry work (Chekroud et al., 2017), for example in predicting the risk of a patient experiencing their first psychotic episode (Koutsouleris et al., 2016; Adams et al., 2016) or categorizing prognosis (Kessler et al., 2016). Other types of predictions include finding best possible treatment for a set of symptoms (Paulus & Thompson, 2019), forecasting treatment response (Webb et al., 2018), or making diagnoses (Kalmady et al., 2019; Hahn et al., 2020). Unsupervised approaches can find new endophenotypes of a disorder that might cause variation in therapeutic responses (Drysdale et al., 2017; Chand et al., 2020).

In the face of overwhelming data, choosing the correct approach is crucial. To develop clinically effective systems, all relevant data should be considered, including electronic health records, -omics, imaging, internet activity, and more. The abundance of data can introduce corrupting noise, which necessitates statistical techniques to enhance signal. Furthermore, contemporary databases tend to skew towards specific populations, such as white men or UK citizens, which must be corrected (Monteith et al., 2015). Finally, data privacy is of utmost concern.

WHY HAS COMPUTATIONAL PSYCHIATRY NOT YET TRANSLATED TO CLINICAL PRACTICE?

In other medicines, computational approaches are flourishing, for example in classifying the presence of cancer (Yoo et al., 2019). Deep learning (see Theoretical Approaches for a brief explanation) to classify diabetic retinopathy has entered clinical trials (Rajalakshmi, 2020). However, psychiatric data are often not as straightforward, given their neurological complexity, subjective nature of experience, and clinical heterogeneity.

While numerous studies have demonstrated the impressive capability of these systems, they maintain insufficient generalizability, such that clinicians have not adopted or even tested them in randomized clinical trials (Woo et al., 2017). While clinical psychiatry itself is imperfect, occasionally prescribing drugs through trial and error, which risks long-term side effects, replacing this with a similarly erroneous system is illogical and expensive (Chekroud & Koutsouleris, 2017). Until computational psychiatry can create useful solutions, it will remain out of the clinic.

It should be noted that the methods themselves are novel. Neuroimaging is both time consuming and expensive (Vu et al., 2018, Chandler et al., 2019). The shift in perspective of fMRI studies from functional region to whole-brain approaches (Richiardi, 2013) emphasizes that a stronger understanding of neural computation is necessary for selecting both methods and relevant data in research. Neuroimaging has proved especially difficult in the search for biomarkers—Dwyer et al. (2018) notes that fMRI studies utilizing classical statistics have a 70% false positive rate, meaning that experiments will find a statistically significant correlation in the data more often than not, even if none is truly there. Similarly, current treatment-predictive models do not incorporate the ability to select multiple therapies, a regular practice in the clinic.

Woo et al. (2016) emphasizes that most computational methods for psychiatry remain in the research stages of development. Yet data-driven approaches have often outmatched clinical counterparts in various clinically-relevant tasks (Bzdok & Meyer-Lindenberg, 2018). So why are these results insufficient to be instantiated in hospitals? It should be noted that most disorders lie in some abstract symptom space where different medical parameters define the dimensions, and an expert-defined decision boundary, which classifies data points based on their location relative to the boundary, determines diagnosis. Samples near the boundary will be difficult to classify, especially if we do not understand the nature of the disease. In these cases, which are frequent, naive approaches might not work as desired, leading to poor generalization among patient types (Schultze-Lutter et al., 2018). Yet some models have found relative success, such as Chekroud et al. (2017) which found three generalizable symptom clusters in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) dataset (MDD, n = 4039) that characterized antidepressant responsiveness.

FAILURE TO REPLICATE

Neuroimaging studies tend to overfit, or fail to generalize beyond, the experimental data. These studies suffer from the curse of dimensionality due to the small sample size. As described by Huys et al. (2016), when the number of features exceeds the number of samples, it is possible to perfectly distinguish n patients from m controls by using n+m-1 features. Functional connectivity matrices, which measure in fMRI how the changes in activity of one brain region correlates with the changes in another, often have close to 100,000 features per data point, while sample sizes are minimal, with close to 100 subjects (Venkatesh et al., 2020). Utilizing a region of interest approach, which focuses on a particular location of the brain, reduces dimensions drastically. Increasing the size of datasets or selecting meaningful features via regularization techniques or theory could further mitigate these issues (Huys et al., 2016).

As mentioned previously, psychiatric data is noisy. Approaches that focus on diagnosis directly from data, such as Zhu et al. (2018), suffer from high rates of misclassifications (Chekroud, 2017) and do not exceed clinical accuracies as they still utilize the DSM-V criteria. Biomarkers might be shared across disorders when these definitions are used, further confounding separability (Fernandes & Berk, 2017).

Commonly used algorithms cannot represent the complex relationships required in psychiatry. Powerful methods such as deep learning have been examined but only as proof-of-concept (Durstewitz et al., 2019; He et al., 2020). Additionally, common optimization techniques such as feature selection, which selects variables based on how much they improve a model, can have detrimental effects when attempting to generalize to new populations (Paulus et al., 2019).

However, the key problem behind data-driven failures is that their methods do not take into account the temporality and plasticity of mental illness. These models capture a snapshot of the clinical picture, abstracting away the dynamics of neurological function. Thus, despite trends towards best practices, they will likely fail to ultimately reach the acceptable threshold of generalizability for clinical usage alone.

CURRENT TRENDS

The two obvious solutions to the major problems (limited data and lack of generalizability) are currently being addressed by open-source projects, triggering an upward trend in sample size. The community has responded via initiatives such as the Human Brain Project and the Human Connectome Project, which have collected large databases of fMRI recordings from thousands of people (Vu et al., 2018). These projects are an excellent step in the right direction, and have yielded significant findings in basic neuroscience research. However, scientists seeking to use the data for clinical research continue to be wary of these sample sizes, as well as the fact that the data regularly comes from one source, which can increase bias (Smith & Nichols 2018). He et al. (2020) found that increasing the sample size on a behavioral and demographic classification task from 100 subjects to 8000 improved the correlation of predicted and ground truth labels from <0.05 to 0.25, a promising increase. Similarly, Hahn et al. (2020) utilized data from 27 recording sites provided by the ENIGMA Addiction working group, which limits single-site bias.

Yet these are often not enough. Drysdale et al. (2017) found two clusters of depression with different symptom profiles based on resting state fMRI that offered separable clinical symptom profiles and differential treatment responses to Transcranial Magnetic Stimulation on 1,188 training samples over multiple sites. Despite these precautions, the study failed to replicate (Chekroud 2020, personal correspondence). If computational methods cannot satisfy the robustness criteria of research, how can we hope to integrate them in practice?

As studies move into clinical testing, they will need even more rigorous standards of validation. Algorithms will need to display reliability (performing adequately for long periods of time), scalability (increasing production and distribution to customers), and ease of implementation (allowing non-experts to utilize the technology) (Nair et al., 2020) via clinical trials. Paulus et al. (2016) detail a prospective pipeline, with phases requiring robustness, clinical validity, efficacy in a randomized clinical trial, clinical effectiveness, and post-marketing refinement. Five years later, no method has passed phase one. What further changes could facilitate progress? To develop a robust predictive or explanatory model of mental health disorders, data should be used in the same way as psychiatrists. Clinicians take the past into account via patient histories, and so too should computational systems (Stiefel et al., 2019). Second, increased emphasis should be placed on theory-based research, as models derived from theory are more likely to generalize and potentially lead to clinically-relevant findings (Huys et al., 2016). While machine learning is effective at tasks such as image recognition (Kirzhevsky et al., 2012), these are not as complex as psychiatry or neuroscience. An informed approach is paramount.

This is not to say that the data-driven research should be abandoned entirely. Rather, it will have a position in clinical practice, perhaps as a first pass system (see my four-step proposal), while the more neuroscientifically-grounded models will further the analysis. Woo et al. (2017) note that the majority (75%) of neuroimaging studies that search for biomarkers for disorder apply a data-driven approach, underscoring the community's excitement towards ML, but excitement is not enough. Similarly, the weak explanatory power of genomics or neuroimaging is not enough to directly prove informative or clinically efficacious (Chekroud, 2018).

THEORETICAL APPROACHES

Theory-driven models draw upon decades of neuroscience (Flagel et al., 2019). They include biophysical simulations and behavioral models in varying degrees of precision. They can account for heterogeneity in standard pathophysiology by adjusting various pieces of a generalized framework to better fit individual subjects (Murray et al., 2018). In the following subsections, I briefly detail a few examples of theoretical models.

Generative Models

Generative models make inferences about unobservable neural states by sequentially taking in data, usually from neuroimaging, and updating the inner state of the model to better match the data. Neuroimaging models are based on properties of functional connectivity networks, which generalize small-scale neural features to systems-level responses (Stephan et al., 2015). fMRI, the predominant form of neuroimaging for these models, measures the Blood-Oxygen Level Dependent (BOLD) response. This is a correlate of neural activity, recorded at a millimeter scale, that abstracts layers of microcircuit interactions and single-neuron physiology. The most popular form of generative models, called dynamic causal models, utilize a system of mathematic differential equations to represent high-level features in fMRI (Stephan et al., 2015). These can be used in psychiatry to examine how different dysfunctions in the neural state can lead to the observations from experimental recordings. In the future, psychiatrists could fit these models to patients to gain a deeper understanding of their specific biological dysfunction.

Reinforcement Learning

Reinforcement learning (RL) models have seen newfound success in representing psychiatric dysfunction. RL is a machine-learning approach that seeks to build adaptive algorithms that can maximize reward in a so-called environment. These are not explicitly taught the solution, as in supervised learning, but rather have to figure it out themselves. Neuroscientific RL is paralleled by artificial intelligence research, and contributions in one domain benefit the other. Computational algorithms such as the successor representation, an efficient form of reinforcement learning that has empirical ties to the function of the striatum in the brain (Dayan, 1993; Gershman, 2018), draw from both neuroscience and artificial intelligence (AI). Huys et al. (2015) used this algorithm to argue that depressive symptoms might draw from dysfunction in state-action evaluation, which is a particular step in the RL framework that requires an agent to choose a particular action given the state of the environment.

Through decision theory, an interdisciplinary field that seeks to study how decisions are made from an algorithmic and statistical perspective, psychiatric disorders can be viewed as occurring via self-reinforcing behavioral dysfunction: solving the wrong problem, such as in substance addiction, solving the right problem in the wrong manner, and solving the right problem correctly, but in the wrong environment, such as post-traumatic stress responses (Huys et al., 2015). These models interpret the effect of Selective Serotonin Reuptake Inhibitors (SSRIs) as normalizing the learning processes, which explains delayed antidepressant response via the corollary that further experience is necessary to relearn healthy behaviors. Some neuroscientists have sought to localize various psychiatric dysfunctions to the cortico-striato-thalamo-cortical loops, which RL connects to deficits in model-based learning or learning algorithms that build models of their environment to play more adaptively (Huys et al., 2016). Biological models can more precisely represent these circuits and have predictive power for disease progression or treatment effects.

Deep Learning

Deep learning has a unique connection to neuroscience as it is based on a reductive model of biological neural networks--neurons are viewed as simple computation devices that gain expressive power through their processing in a parallel and distributed manner--and can therefore model neural systems. Image recognition networks have been shown to replicate the visual cortex functional hierarchy (Richards et al., 2019). These have not yet been used to explicitly model psychiatric dysfunction, but initial forays should be expected in the near future. These systems are also effective in data-driven algorithms, and it is plausible that they will be used in each manner.

Hybrid Models

Hybrid models seek to utilize theory to develop features for data-driven models, which can improve predictive power by reducing noise in the data. Brodersen et al. (2011) modelled auditory cortex functional connectivity to identify aphasics--people who have lost the capacity of speech due to brain damage--with 98% accuracy. In this system, various generative models are fit to a dataset, followed by application of supervised algorithms on the features in the generative models (Brodersen et al., 2014; Stephan et al., 2015; Wicki et al., 2015). With a hybrid model, Frassle et al. (2020) classified depressive patients as chronic versus remissive with 79% accuracy, although the training set was quite small, at 85 subjects. Similar to deep learning models, more of these approaches should be expected in the near future. Using these models might be the most effective single way to bring computational psychiatry into the clinic, as it leverages the benefits of each type of approach.

Theoretical models lean more heavily on the research side of computational psychiatry, and therefore they have been treated with skepticism as to their potential efficacy in a clinical setting. However, a properly designed model provides a general framework that can be tailored to an individual patient, thus allowing for precision medicine, much like a laboratory test provides specific measurements that can be tied to a theoretical model of physiology to gain insight into that particular patient's disorder.

COMPUTATIONAL APPROACHES ON CLINICAL DEPRESSION

In this section, I review a selection of studies on Major Depressive Disorder. These are not exhaustive but indicative of current trends.

Data-Driven Studies

Patel et al. (2016) summarize early computational psychiatry studies that use MRI data and focused on diagnosis. None of the patient samples exceeded 80 subjects, and methods tended to be linear, usually filtering voxels--individual pixels in an fMRI recording--with an unsupervised algorithm or functional knowledge. The following studies utilize more modern approaches.

Islam et al. (2018) extracted data from 7,145 Facebook comments to identify phrases that could predict depression, which they identified via a supervised model. They identified phrases with emotional, temporal, social, or perceptual qualities that significantly predicted onset of MDD. Chekroud et al. (2018) similarly used a dataset with 20,785 subjects from U.S national surveys to determine whether a patient would seek treatment, doing so with 70.6% accuracy. It should be noted that this dataset was skewed female and white (72% and 77% respectively). Relevant predictors for initiation of treatment included dropping out of college or having no serious suicidal ideation. This model exemplifies how the computational methods discussed in this paper can not only prove relevant for clinical research, but also for a clinical setting. One would simply have to input phrases that a patient used into this system to determine whether or not they might be depressive.

Webb et al. (2018) identified a subset of 216 MDD patients that preferentially responded to sertraline (an SSRI) who were older, employed, more neurotic and depressive, and having stronger cognitive control than average. Bondar et al. (2020) utilized an unsupervised learning algorithm to identify two symptom clusters in adolescent depressives (n = 439), in which the first (social withdrawal, insomnia, fatigue, etc.) responded well to fluoxetine, an SSRI, and cognitive behavioral therapy, whereas the other (increased appetite, guilt, suicidal ideation, etc.) did not. Chekroud et al. (2016) utilized the open-source STAR*D database to identify variables to predict remission after citalopram treatment, finding significant contribution from employment status, psychomotor agitation, race, education, and more. Importantly, these are features that a psychiatrist might deem relevant.

Theoretical Studies

Generative models of depression are especially difficult to develop due to the heterogeneity of the disorder. Depression is associated with deficits in reward learning, especially in effort valuation (Husain & Roiser, 2018). Kumar et al. (2008) localized diminished prediction error signals in the ventral striatum, which correlated with a reduction in responsiveness to antidepressants. However, Rutledge et al. (2017) disputed this result in a larger sample, finding that moderately depressed patients maintained control-level reward prediction error signals. They utilized a computational model of happiness and found that severe MDD patients fit to this model differed only from controls by a static mood intercept, which the authors interpreted as a dysfunction in higher-order processing. These results agree with the psychological theory of baselines, which argues that a person's happiness at any given time is related to their baseline quality of life (Young et al., 1996).

HOW ARE THEORY-DRIVEN MODELS BETTER SUITED TO NEUROPSYCHIATRY?

Theory-driven models emphasize underlying neural pathology. Biologically-driven theories attempt to explain features, such as the dysfunction of neurotransmitter systems, that can be further represented mathematically (Stephan et al., 2015). These models have strong predictive capabilities and can be further validated in translational animal studies, which allow for invasive experiments (Stephan et al., 2015). Additionally, they allow for the simulation of realistic data which could be used to predict disease progression (Frassle et al., 2017). Data-driven approaches do not have these capabilities, and these "black-box" models--so-called because their inner workings are not fully understood--can learn discriminatory representations if the data itself is biased, a historical problem in medicine. On the other hand, the interpretability and strict assumptions of theoretical models limit bias (Rutledge et al., 2019; Chandler et al., 2019).

Generative models that holistically represent dysfunctions as parameters or dynamics can be directly connected with individual patients, thus "treating the patient not the disease" (Stephan et al., 2015). Such approaches can additionally account for fine-grained changes which ripple to the global scale and to behavior. Because psychiatric disorders have been associated with systemic neuromodulator dysfunction, this is appealing. For example, the dopaminergic system is hypothesized to function as a prediction error signal, which is the learning signal in RL (Schultz et al., 1997). Similarly, serotonin has been theorized as a discounting parameter in a utility function, although it certainly has multiple functions (Huys et al., 2015). A discounting parameter is another feature in an RL algorithm that quantifies how much an agent "cares" about the future relative to the present, which is measured as an exponentially weighted sum of expected rewards. These theories provide explanations for the effects of therapeutics, while data-driven approaches cannot. Even better, generative models of adaptive plasticity can predict the mechanisms of treatment response based on the patient's "neurotype" (Vinogradov, 2017).

Frassle et al. (2017) note that high dimensionality—the sheer number of features per data point—of neuroimaging introduces high levels of variance that is challenging even for ML. However, biologically interpretable features can separate classes of patients, on which traditional ML techniques can make predictions. An added benefit is that models can be compared to optimally explain a set of symptoms (Bennett et al., 2019).

WHAT ARE SOME ISSUES WITH THEORY-DRIVEN MODELS?

Precise mechanistic models are needed to sufficiently capture neural dynamics, which is a huge challenge. An incorrect or non-parsimonious model (one that is not sufficiently simplified while remaining precise) is likely to extract results from noise and therefore overfit (Deco & Kringelbach, 2014). The temporospatial restrictions of neuroimaging, as mentioned above, limit the ability of generative models to represent underlying activity. The complexity of whole brain models makes optimization increasingly intractable (not computable in a reasonable amount of time). Ultimately, an increase in computational power, sample size, and algorithmic heuristics will be required to train these systems to a functional level, just as the deep learning community found in the early 2010s (Chen & Lin, 2014). Theory-driven approaches have mainly focused on schizophrenia to date, but future trends will include other disorders (deFilippis et al., 2019). Like data-driven models, these systems have yet to move past the exploratory phase (Frassle et al., 2017).

Despite the fact that these methods are still in their infancy, it is likely that just as in other medicines, psychiatrists will soon implement artificial intelligence to aid their decision making. How might this look in practice? In the following section, I provide a novel four-step proposal that seeks to use the computational tools, developed by contemporary and future research, in a maximally-effective manner to treat mental health disorders.

A NOVEL PROPOSAL FOR THE CLINICAL IMPLEMENTATION OF COMPUTATIONAL SYSTEMS

In this section, I envision an integration of theory and data-driven models in clinical practice. The proposal contains four basic steps with a recurrence paradigm for long-term treatment when computational psychiatry approaches are sufficient for medicinal use. The following steps require a comprehensive set of algorithms to utilize all informative data. Note that this is a general plan and would require further personalization for precision medicine.

1. Immediate Treatment

Many psychiatric disorders require immediate treatment, such as suicidal ideation. Data-driven algorithms using information that is immediately collectable can provide initial treatment recommendations.

2. Biological data and theoretical models

Many psychiatric disorders require immediate treatment, such as suicidal ideation. Data-driven algorithms using information that is immediately collectable can provide initial treatment recommendations.

3. Longitudinal data collection

Over a specified time, the patient utilizes a smartphone application to record relevant data, such as sleep and movement, alongside surveys or virtual therapy sessions. These are factored into the history portion of the model in order to capture the dynamics of the patient's disorder.

4. Informed, holistic treatment

As treatment continues, the historical information is integrated into a single, cumulative model, and a clinician designs a more general treatment plan. Efficacy of the treatment can be revised by repeating these steps. This precision medicine approach accounts for many of the elements of experience desired by vocal opponents to personalized psychiatry (Stiefel at al., 2019). No single step will be sufficient, as indicated by preliminary research.

DISCUSSION

The development of computational psychiatry is still exploratory; clinical efficacy is far off. ML is a necessary tool but not a silver bullet; applying these models unintelligently will not suddenly solve decades-old problems. Simon (2019) makes an excellent analogy, emphasizing that despite the hype of ML, we cannot become like a child with a hammer, pounding anything that looks like a nail. Contrarians argue that computational models cannot be as effective as a clinician, because they do not have an understanding of subjective experience.

Translational computational tools need to derive from basic science. Neuroscience and psychiatry will benefit greatly from scientists who have rigorously studied theory and methodology (Mai-an Vu et al., 2018; Cearns et al., 2019). Theoretical approaches have yet to begin answering the questions desired of computational psychiatry due to extensive methodological development. Asking the right questions is crucial, and we must take the time to do so. Are our computational tools powerful enough for these approaches? The answer is yes, but the more pertinent question is whether we have the right type of data. Computational psychiatry researchers are hence cautiously optimistic about the clinical viability of ML methods (Chekroud & Koutsouleris, 2017).

CONCLUSION

Perhaps it is too early to determine whether theoretical or data-driven approaches will be more efficacious for the future of computational psychiatry and clinical practice. In all likelihood, both methods will be necessary. The majority of research in this field requires a stronger theoretical foundation that will currently hinder the development of clinical tools, but it is still important to consider how clinical research can translate. This will be useful to psychiatry in general, as biologically-backed theories can help improve the definitions and treatments of disorders in the DSM. Psychiatritfsts will of course never be phased out, but machine learning algorithms can pick up trends that even the expert eye cannot capture in vast amounts of data. Furthermore, efforts to create testable theoretical models must keep pace with their counterpart as these studies will be more informative in the long run. Scientists, hospitals, and therapy developers will need to communicate intensively to steer psychiatry into a new era. With time, psychiatry will soon join other disciplines in the era of precision medicine.

ACKNOWLEDGMENTS

The author thanks Professor Avram Holmes for rigorous discussion and guidance both during the development of this project and afterwards.

REFERENCES

Adams, R. A., et al. (2016). "Computational Psychiatry: towards a mathematically informed understanding of mental illness." Journal of Neurology, Neurosurgery & Psychiatry 87(1): 53.

Bennett, D., et al. (2019). "The Two Cultures of Computational Psychiatry." JAMA Psychiatry 76(6): 563-564.

Bondar, J., et al. (2020). "Symptom clusters in adolescent depression and differential response to treatment: a secondary analysis of the Treatment for Adolescents with Depression Study randomised trial." The Lancet Psychiatry 7(4): 337-343.

Brady, E. U. and P. C. Kendall (1992). "Comorbidity of anxiety and depression in children and adolescents." Psychological Bulletin 111(2): 244-255.

Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.

D., Buhmann, J. M., & Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. NeuroImage: Clinical, 4, 98-111.

Brodersen, K. H., et al. (2011). "Generative Embedding for Model-Based Classification of fMRI Data." PLOS Computational Biology 7(6): e1002079.

Browne, C. and I. Lucki (2013). "Antidepressant effects of ketamine: mechanisms underlying fast-acting novel antidepressants." Frontiers in Pharmacology 4(161).

Bzdok, D. and A. Meyer-Lindenberg (2018). "Machine Learning for Precision Psychiatry: Opportunities and Challenges." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 3(3): 223-230.

Cearns, M., et al. (2019). "Recommendations and future directions for supervised machine learning in psychiatry." Translational Psychiatry 9(1): 271.

Chand, G. B., Dwyer, D. B., Erus, G., Sotiras, A., Varol, E., Srinivasan, D., ... & Davatzikos, C. (2020). Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain, 143(3), 1027-1038.

Chandler, C., et al. (2019). "Using Machine Learning in Psychiatry: The Need to Establish a Framework That Nurtures Trustworthiness." Schizophrenia Bulletin 46(1): 11-14.
Chekroud, A. M. (2017). "Bigger Data, Harder Questions—Opportunities Throughout Mental Health Care." JAMA Psychiatry 74(12): 1183-1184.

Chekroud, A. M., et al. (2018). "Predicting Barriers to Treatment for Depression in a U.S. National Sample: A Cross-Sectional, Proof-of-Concept Study." Psychiatric Services 69(8): 927-934.

Chekroud, A. M., et al. (2017). "Reevaluating the Efficacy and Predictability of Antidepressant Treatments: A Symptom Clustering Approach." JAMA Psychiatry 74(4): 370-378.

Chekroud, A. M., et al. (2016). "Cross-trial prediction of treatment outcome in depression: a machine learning approach." The Lancet Psychiatry 3(3): 243-250.

Chen, X. W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. IEEE access, 2, 514-525.

Cho, G., et al. (2019). "Review of Machine Learning Algorithms for Diagnosing Mental Illness." Psychiatry investigation 16(4): 262-269.

Dayan, P. (1993). "Improving Generalization for Temporal Difference Learning: The Successor Representation." Neural Computation 5(4): 613-624.

de Filippis, R., Carbone, E. A., Gaetano, R., Bruni, A., Pugliese, V., Segura-Garcia, C., & De Fazio, P. (2019). Machine learning techniques in a structural and functional MRI diagnostic approach in schizophrenia: a systematic review. Neuropsychiatric disease and treatment, 15, 1605.

Deco, G. and Morten L. Kringelbach (2014). "Great Expectations: Using Whole-Brain Computational Connectomics for Understanding Neuropsychiatric Disorders." Neuron 84(5): 892-905.

Drysdale, A. T., et al. (2017). "Resting-state connectivity biomarkers define neurophysiological subtypes of depression." Nature Medicine 23(1): 28-38.

Durstewitz, D., et al. (2019). "Deep neural networks in psychiatry." Molecular Psychiatry 24(11): 1583-1598.

Dwyer, D. B., et al. (2018). "Machine Learning Approaches for Clinical Psychology and Psychiatry." Annual Review of Clinical Psychology 14(1): 91-118.

Fernandes, B. S., & Berk, M. (2017). Staging in bipolar disorder: one step closer to precision psychiatry. Brazilian Journal of Psychiatry, 39(2), 88-89.

Flagel, S. B., et al. (2019). "Editorial: bridging the gap with computational and translational psychopharmacology." Psychopharmacology 236(8): 2291-2294.

Frässle, S., Marquand, A. F., Schmaal, L., Dinga, R., Veltman, D. J., Van der Wee, N. J., ... & Stephan, K. E. (2020). Predicting individual clinical trajectories of depression with generative embedding. NeuroImage: Clinical, 26, 102213.

Frässle, S., et al. (2018). "Generative models for clinical applications in computational psychiatry." WIREs Cognitive Science 9(3): e1460.

Friston, K. J. (2004). "Models of Brain Function in Neuroimaging." Annual Review of Psychology 56(1): 57-87.

Friston, K. J., Redish, A. D., & Gordon, J. A. (2017). Computational nosology and precision psychiatry. Computational Psychiatry, 1, 2-23.

Gershman, S. J. (2018). "The Successor Representation: Its Computational Logic and Neural Substrates." The Journal of Neuroscience 38(33): 7193.

Gillan, C. M., & Whelan, R. (2017). What big data can do for treatment in psychiatry. Current Opinion in Behavioral Sciences, 18, 34-42.

Gordon, J. A. (2010). "Testing the glutamate hypothesis of schizophrenia." Nature Neuroscience 13(1): 2-4.

Hahn, S., Mackey, S., Cousijn, J., Foxe, J. J., Heinz, A., Hester, R., ... & Garavan, H. (2020). Predicting alcohol dependence from multisite brain structural measures. Human Brain Mapping.

He, T., Kong, R., Holmes, A. J., Nguyen, M., Sabuncu, M. R., Eickhoff, S. B., ... & Yeo, B. T. (2020). Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. NeuroImage, 206, 116276.

Husain, M. and J. P. Roiser (2018). "Neuroscience of apathy and anhedonia: a transdiagnostic approach." Nature Reviews Neuroscience 19(8): 470-484.

Huys, Q. (2013). Computational Psychiatry. Encyclopedia of Computational Neuroscience. D. Jaeger and R. Jung. New York, NY, Springer New York: 1-10.

Huys, Q. J. M., et al. (2015). "Depression: A Decision-Theoretic Analysis." Annual Review of Neuroscience 38(1): 1-23.

Huys, Q. J. M., et al. (2015). "Decision-Theoretic Psychiatry." Clinical Psychological Science 3(3): 400-421.

Huys, Q. J. M., et al. (2016). "Computational psychiatry as a bridge from neuroscience to clinical applications." Nature Neuroscience 19(3): 404-413.

Insel, T., et al. (2010). "Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders." American Journal of Psychiatry 167(7): 748-751.

Islam, M. R., et al. (2018). "Depression detection from social network data using machine learning techniques." Health information science and systems 6(1): 8-8.

Jahanshad, N. (2019). "14.1 Large-Scale Psychiatric Imaging Studies in Enigma: What is Gained and What are Future Challenges." Schizophrenia Bulletin 45(Supplement_2): S110-S111.

Kalmady, S. V., et al. (2019). "Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning." npj Schizophrenia 5(1): 2.

Kessler, R. C., et al. (2016). "Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports." Molecular Psychiatry 21(10): 1366-1371.

Koutsouleris, N., et al. (2016). "Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach." The Lancet Psychiatry 3(10): 935-946.

Krizhevsky, A., et al. (2012). ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Lake Tahoe, Nevada, Curran Associates Inc.: 1097–1105.

Kumar, P., et al. (2008). "Abnormal temporal difference reward-learning signals in major depression." Brain 131(8): 2084-2093.

Liu, G. D., Li, Y. C., Zhang, W., & Zhang, L. (2020). A brief review of artificial intelligence applications and algorithms for psychiatric disorders. Engineering, 6(4), 462-467.

Lynch, L. K., et al. (2018). "Task-evoked functional connectivity does not explain functional connectivity differences between rest and task conditions." Human Brain Mapping 39(12): 4939-4948.

Marblestone, A. H., et al. (2016). "Toward an Integration of Deep Learning and Neuroscience." Frontiers in Computational Neuroscience 10(94).

Mitchell, A. J. (2018). "Two-week delay in onset of action of antidepressants: new evidence." British Journal of Psychiatry 188(2): 105-106.

Montague, P. R., et al. (2012). "Computational psychiatry." Trends in Cognitive Sciences 16(1): 72-80.

Monteith, S., et al. (2015). "Big data are coming to psychiatry: a general introduction." International Journal of Bipolar Disorders 3(1): 21.

Murray, J. D., Demirtaş, M., & Anticevic, A. (2018). Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3(9), 777-787.

Nguyen, K. P., Fatt, C. C., Treacher, A., Mellema, C., Trivedi, M. H., & Montillo, A. (2020, March). Anatomically informed data augmentation for functional MRI with applications to deep learning. In Medical Imaging 2020: Image Processing (Vol. 11313, p. 113130T). International Society for Optics and Photonics.

Orrù, G., et al. (2012). "Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review." Neuroscience & Biobehavioral Reviews 36(4): 1140-1152.

Patel, M. J., Khalaf, A., & Aizenstein, H. J. (2016). Studying depression using imaging and machine learning methods. NeuroImage: Clinical, 10, 115-123.

Paulus, M. P., Huys, Q. J., & Maia, T. V. (2016). A roadmap for the development of applied computational psychiatry. Biological psychiatry: cognitive neuroscience and neuroimaging, 1(5), 386-392.

Paulus, M. P., et al. (2019). "Machine Learning and Brain Imaging: Opportunities and Challenges." Trends in Neurosciences 42(10): 659-661.

Paulus, M. P. and W. K. Thompson (2019). "The Challenges and Opportunities of Small Effects: The New Normal in Academic Psychiatry." JAMA Psychiatry 76(4): 353-354.

Paulus, M. P., & Thompson, W. K. (2019). Computational approaches and machine learning for individual-level treatment predictions. Psychopharmacology, 1-9.

Poldrack, R. A. and K. J. Gorgolewski (2014). "Making big data open: data sharing in neuroimaging." Nature Neuroscience 17(11): 1510-1517.

Rajalakshmi, R. (2020). "The impact of artificial intelligence in screening for diabetic retinopathy in India." Eye 34(3): 420-421.

Richiardi, J., et al. (2013). "Machine Learning with Brain Graphs: Predictive Modeling Approaches for Functional Imaging in Systems Neuroscience." IEEE Signal Processing Magazine 30(3): 58-70.

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. Current opinion in neurobiology, 55, 152-159.

Rutledge, R. B., et al. (2017). "Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression." JAMA Psychiatry 74(8): 790-797.

Schultze-Lutter, F., Schmidt, S. J., & Theodoridou, A. (2018). Psychopathology—a precision tool in need of re-sharpening. Frontiers in psychiatry, 9, 446.

Seeman, P. (1987). "Dopamine receptors and the dopamine hypothesis of schizophrenia." Synapse 1(2): 133-152.

Shatte, A. B. R., et al. (2019). "Machine learning in mental health: a scoping review of methods and applications." Psychol Med 49(9): 1426-1448.

Simon, G. E. (2019). "Why the Nails Should Boss the Hammers." Psychiatric Services 70(8): 642-643.

Smith, S. M. and T. E. Nichols (2018). "Statistical Challenges in "Big Data" Human Neuroimaging." Neuron 97(2): 263-268.

Steele, J. D. and M. P. Paulus (2019). "Pragmatic neuroscience for clinical psychiatry." The British Journal of Psychiatry 215(1): 404-408.

Stephan, Klaas E., et al. (2015). "Translational Perspectives for Computational Neuroimaging." Neuron 87(4): 716-732.

Stewart, R. and K. Davis (2016). "'Big data' in mental health research: current status and emerging possibilities." Social Psychiatry and Psychiatric Epidemiology 51(8): 1055-1072.

Stiefel, F., et al. (2019). "Precision psychiatry: Promises made—Promises to be kept?" Australian & New Zealand Journal of Psychiatry 53(9): 841-843.

Turner, B. O., et al. (2018). "Small sample sizes reduce the replicability of task-based fMRI studies." Communications Biology 1(1): 62.

Venkatesh, M., Jaja, J., & Pessoa, L. (2020). Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. NeuroImage, 207, 116398.

Vu, M.-A. T., et al. (2018). "A Shared Vision for Machine Learning in Neuroscience." The Journal of Neuroscience 38(7): 1601-1607.

Wang, X.-J. and John H. Krystal (2014). "Computational Psychiatry." Neuron 84(3): 638-654.

Webb, C. A., et al. (2018). "Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study." Psychological Medicine 49(7): 1118-1127.

Whitfield-Gabrieli, S., et al. (2016). "Brain connectomics predict response to treatment in social anxiety disorder." Molecular Psychiatry 21(5): 680-685.

Wiecki, T. V., et al. (2015). "Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification." Clinical Psychological Science 3(3): 378-399.

Woo, C.-W., et al. (2017). "Building better biomarkers: brain models in translational neuroimaging." Nature Neuroscience 20(3): 365-377.

Yang, Z., et al. (2020). "Unsupervised Classifications of Depression Levels Based on Machine Learning Algorithms Perform Well as Compared to Traditional Norm-Based Classifications." Frontiers in Psychiatry 11(45).

Yoo, S., et al. (2019). "Prostate Cancer Detection using Deep Convolutional Neural Networks." Scientific Reports 9(1): 19518.

Young, M. A., et al. (1996). "Stable trait components of hopelessness: Baseline and sensitivity to depression." Journal of Abnormal Psychology 105(2): 155-165.

Zeng, L.-L., et al. (2014). "Unsupervised classification of major depression using functional connectivity MRI." Human Brain Mapping 35(4): 1630-1641.

Zhou, Z., et al. (2020). "Machine learning methods in psychiatry: a brief introduction." General psychiatry 33(1): e100171-e100171.

ABOUT THE AUTHOR

Mitchell Ostrow, Pauli Murray '22

by Matthew Fan, Benjamin Franklin '24

In the Seo Lab at the Yale School of Medicine, senior Mitchell Ostrow found his passion at the intersection of computational modeling, machine learning, and neuroscience—studying deep neural networks as models of the brain. To Ostrow, studying these models is especially exciting because through artificial intelligence, his findings can directly impact the world on top of moving science forward. For example, they could potentially be used to synthesize drugs or devise new treatments immediately. Ultimately, Ostrow's commitment to pursuing what intrigued him the most led him to this research area.

"From doing so much exploration, I was able to really narrow down my interests and find something that I absolutely love and can definitely see myself doing for the rest of my life," Ostrow said. Right now, that means pursuing a PhD in Computational Neuroscience to study the intersection of AI and neuroscience.

Outside of research, Ostrow enjoys exercise and spending time in nature. Additionally, he is heavily invested in music, formerly playing trombone in the Yale Symphony Orchestra, a trombone choir called Scale and Bones (which he founded), and a brass choir called Coup de Brass. Before college, his identity was predominantly as a trombone player and as a musician. Although he still sees himself in this way, his identity has transformed into that of a researcher.

Surprisingly, he has found commonalities among these two worlds. Initially, most of your time is spent developing technical skills—such as playing scales for trombone and learning how to analyze papers in research. As you progress, however, you develop your own style or you create your own experiments, and creativity flourishes. Moreover, music and research are both personally rewarding as well as community-oriented.

"Science is for society to gain knowledge and music is for other people to enjoy," Ostrow said. "To me, it's more about appreciating the music or appreciating creating knowledge for myself, and it's an added benefit that other people enjoy it."

For the full-length profile, visit yalesymposia.com