Proposal Summary Slide

A Complexity Theory of Physiology

Complexity Working Group; POC & PI: Daniel Ari Friedman Ph.D.

HR001118S0058 (12) Polyplexus Pilot 2
Proposal Summary Slide

Concept

Moving beyond snapshots, simple parametric descriptives, and single sensors.

Leveraging multiple data streams to search for "Big Mechanisms" and causal relationships in physiological processes.

Impact

- Provide valuable insight on disease prevention and health cost savings.
- Provide foundation for follow-up studies regarding instantaneous physiological intervention in the field via wearables.
- Reveal principles and mechanisms of analogous complex systems.
- Shift perspectives on health and data, from symptomatology and correlates to causation and auditability.

Approach

Time Series Dynamics vs Snapshots
"Big Mechanisms" vs "Big Data"

- Literature review on biological frameworks for physiological integration & causal inference models.
- Move away from unsupervised, pure-data driven models, which find correlation but not causation, and move towards mechanism-based approaches.
- Move away from simple, and often expensive snapshot measures and move towards passively and continuously collected measures from a variety of wearable devices.

Synopsis

Complexity Theory is an interdisciplinary approach for the study of dynamic systems that are composed of interacting subunits. Organisms display many of the hallmarks of complex adaptive systems, such as functional unity from many coordinating subunits, non-linear responsivity to ecological stimuli, and fractal scaling across spatio-temporal scale. Decades of research into physiological systems has revealed basic principles of bodily function (e.g. homeostasis, hormesis, learning, robustness), but these generalized system properties are often difficult to apply in the specifics. Within the last 5 years, the commercial market has been flooded with new kinds of wearable devices that track physiological metrics including heart rate, breathing rate, and neural signals such as EEG/ECG. These wearables can be used for basic monitoring of health, or for biofeedback in some limited contexts. We believe that by taking Complexity approach to integrating multiple sensors, novel and powerful insights into human

physiology will be obtained.

The novelty of our approach, described below, arises from the manner in which we are contextualizing sensor data. Rather than simply fitting descriptive mathematical models to physiological data (e.g. by taking the average heart rate, or even the variability of the heart rate as other Complexity approaches have done), we are constructing a generative model of human physiology that is then improved in a personalized fashion via the emissions of physiological data. There are multiple key advantages of combining a Complexity perspective with a generative model. First, we will able to easily deal with incomplete, noisy, or limited sensor data - because we are improving our estimate about system function instead of reinventing the wheel each time, our model will be able to provide insights into personalized health by drawing on impersonal resources (e.g. biological databases of physiological relations, literature on classification of bodily states). From the generative model we will be able to determine which experiments or measurements would be optimally informative given a particular case - something a purely descriptive model will never be able to do. Key technical challenges to the implementation of a generative physiological model include the formidable computational effort of making a coarse-grained model of human physiology, and also integrating multiple heterogeneous sensor data formats in a streaming manner. Here we will take to heart the Complexity maxim that “More is Different”, seeking to find elegant approaches that combine simple information across sensor streams given what we already know about body physiology (e.g. heart rate variability & breathing variability are connected, and related to exercise), rather than trying to use machine learning to find a needle in a haystack.

The deliverables of our project include peer-reviewed/conference papers, and also a flexible software platform for human physiology. The publications related to Complexity and Physiology will rise above the current state of play, as most work in this area is focused on a single metric / single disease (e.g. gait variability & fall risk) rather than a general perspective. Additionally our work would be the first Complexity Physiology research that provides a complex generative model with physiological emissions (similar to Hidden Markov Model, where the hidden states are true body states & physiological variables are observed emissions) instead of simple descriptive statistics. Our publications in the computational realm will also raise the level of work in the area, as we will focus on the implementation of a rigorous theoretical model that explicitly allows for modular addition of more sensors, instead of overfitting single-sensor data to a single use case (e.g. EEG just for meditation). Our work will also allow for many future directions, such as the ability to perform counterfactual analysis on the generative model (e.g. how would the body respond to intervention X), increased ability to incorporate snapshot physiological measures such as blood tests, and the ability to incorporate non-classical physiological measures such as digital device response time or automated detection of posture via video analysis.

Goals

Here we outline the three primary Goals of our proposal to integrate Complexity Science with Physiology through the use of modern computational methods and commercial hardware.

Goal 1: Towards a Complexity Theory and Framework for Physiology.
Goal 2: Big Mechanism & Big (Sensor) Data for Complexity Physiology.
Goal 3: Physiological case studies: Inflammation, Aging, & Learning.

Goal 1: Towards a Complexity Theory and Framework for Physiology.

We propose to develop an integrated framework of Complexity Physiology that combines biological knowledge structures (e.g. ontologies & controlled vocabularies of signaling pathways) with modern computational approaches for understanding and controlling multilevel dynamic systems. We intend to combine the strengths of multiple Complexity tools, some of which have been previously applied to physiology and some of which have not. Tools from Complexity that are already being applied to the study of physiology include fractal analysis of time series data and use of statistics based upon information theory rather than naive parametric models. We intend to differ from, and build off of, this previous work in two main ways. First, we will combine dynamic time series models with formal knowledge models that incorporate heterogeneous knowledge databases. Second, our model will be structured quite differently than previous approaches to physiology – we will implement a generative model that "emits" physiological data, allowing us to combine personalized health data with impersonal knowledge sources such as information about neural rhythms and metabolic pathways.

Goal 2: Big Mechanism & Big (Sensor) Data for Complexity Physiology.

We will apply our Complexity Physiology framework (Goal 1) to human physiology using both custom and commercially-available tools. Human physiology provides a wealth of data that are important for health and performance, including but not limited to EEG, ECG, heart rate, blood pressure/oxygenation/glucose, and movement data. Increasingly we are able to capture these biomarkers in high-resolution longitudinal time series datasets by building upon the technological advancements in the consumer wearables markets. Using these rich personalized datasets, we will apply our developed theoretical models of complexity and explore their fit and predictive power, rapidly refining our models and spurring new theoretical insights in a virtuous cycle of discovery. Part of this effort will be developing new tools to improve both descriptives, including new visualization aids (e.g. dashboards for biofeedback or assessment), and statistical models (e.g. variational methods for using censored data, methods for strong data privacy), building upon probabilistic programming and machine learning advances. With these refinements to our model and tools, we will be prepared to apply our insights to the high impact areas of human health outlined in Goal 3.

Goal 3: Physiological case studies: Inflammation, Aging, & Learning.

After refining our theoretical models on complex physiology datasets representing healthy humans at baseline, we will apply our model and tools to high impact areas of human health. Many emergent pathologies are complex in nature but require onerous blood tests to track. One example we will be targeting is low-grade inflammation associated with age, dubbed ‘inflammaging’, whose dynamics rooted in immune system function is poorly understood. There are several biomarkers that are correlated with low-grade inflammation (e.g. heart rate variability, sleep quality, glucose, and activity levels) that have become readily accessible through new wearable technology. Blood tests exist; however this discrete measurement is cost-prohibitive and not impervious to the inherent variability in this chronic condition. A non-invasive means to measure the cytokine profile of individuals suffering from this phenomenon (e.g. using only wearable sensors) could lead to new discoveries and research avenues that shed light on the mechanisms at work. No mechanistic theory that unites these macro-level biomarkers has been proposed and no research studies have been conducted to evaluate the predictive ability of these biomarkers on inflammation. Accessing data from non-sensor data sources, such as regular blood tests, presents new opportunities (integrating across physiological systems) as well as new challenges (e.g. cost, participant involvement).

Technical Plan – Goal 1

As physiology is a non-linear multiscale dynamical process, various simple parametric descriptors (e.g. moving average, standard deviation) may fail to capture essential systemic information. Recognition of this inadequacy has led to various threads of work related to Complexity Physiology [1–3]. Here we will briefly summarize some of the key relevant work in Complexity Physiology in the context of the major technical challenges in the field. The outcome of these technical challenges is that a tremendous gap exists between theory and practice in physiology: we still mostly are unable to understand, predict, or control physiological states in a quantitative fashion. Here we are oriented toward using recent developments in Complexity Science and computational methods to address two key technical challenges in Physiology.

1. **Physiological systems are sensitive, dynamic, non-linear, and chaotic**, leading to challenges related to classification, prediction, and control of physiological states.
2. **Physiological data are multiscale, complicated, heterogeneous, and noisy**, leading to practical issues when trying to deploy sensors in biofeedback or diagnostic scenarios.

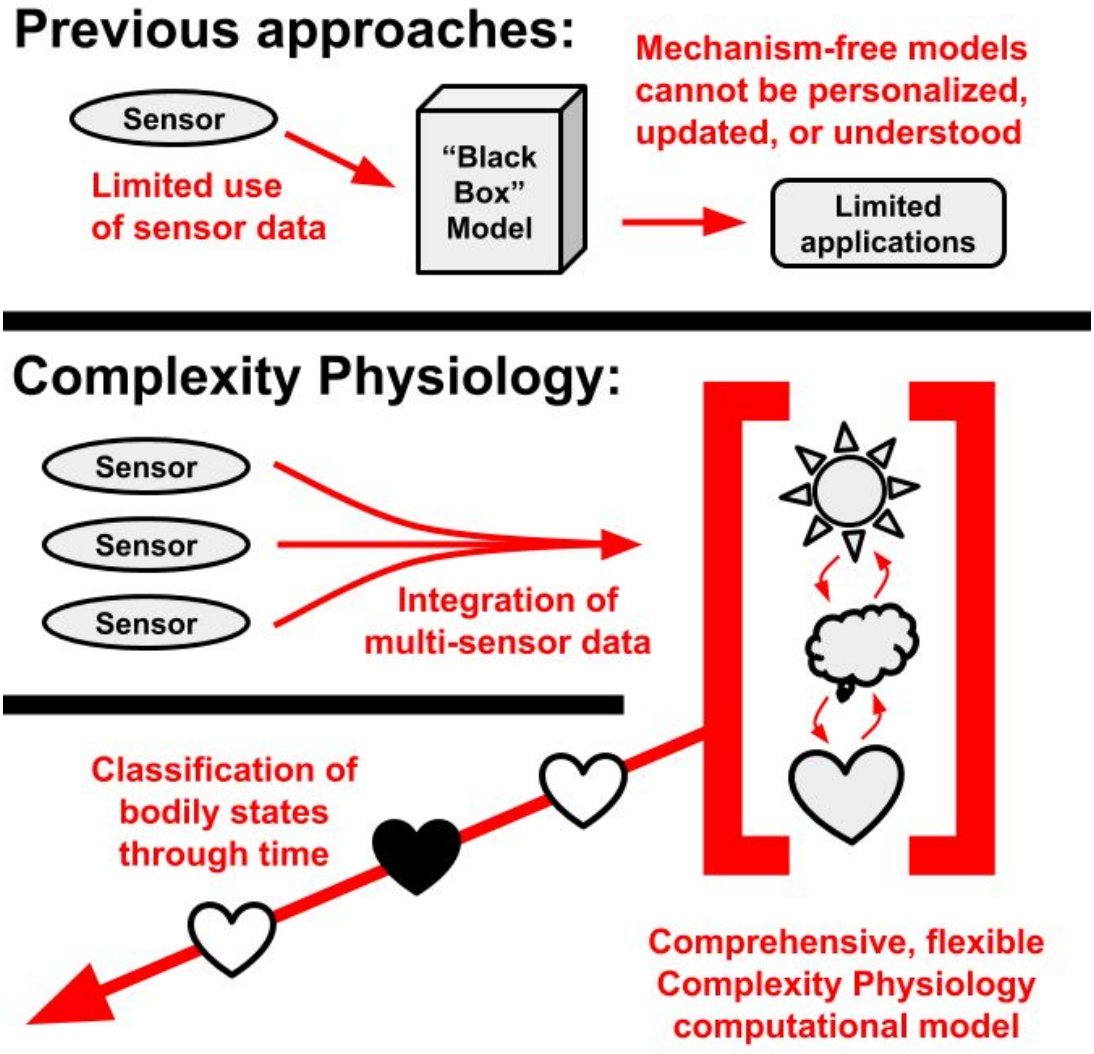
This gap in many ways is barely addressed by research on animal models (flies, mice, etc), as non-human species differ from humans in even essential processes such as wound healing. Additionally non-human species cannot begin to serve as models for understanding the relationship between physiology and human-specific traits, such as the use of digital extended cognitive devices.

Key differences in our approach:

Rather than fit descriptive models to complicated time series data, **we will use longitudinal time series data to reduce our uncertainty about an underlying generative model of human physiology** that will be augmented through multiple channels (e.g. by drawing on findings from literature, environmental datasets, etc). To prevent over-simplification or over-fitting of the underlying generative model of physiology, we will use modern techniques related to hierarchical Bayesian variational methods on complex adaptive systems [4–6]. Briefly, the advantage of the generative approach is that we will be able to optimally finesse the tradeoff between model fit and generalizability by considering raw data (e.g. from a wearable sensor) as a statistical emission from a system that we already know a considerable amount about (the human body). In other words, instead of trying to simply squeeze additional useful descriptive information from single time series data sources (e.g. heart rate variability [7–9]), we will be able to pool the mutual information across multiple physiological data streams *and augment this information* through the use of formal knowledge modeling approaches from Complexity Science and Systems Biology (e.g. meta-conceptual approaches of [10], causal entropic frameworks [11,12] , ontological systems [13,14] , curated biological databases).

Additionally, several other tools from Complexity may be able to be brought to bear on unique aspects of physiological processes. Body physiology is “chaotic” in that the system is (non-linearly) sensitive to small changes – for example the difference between two words might only directly influence a few retina photoreceptors, but the behavioral implications might be

massive. Complexity Science has demonstrated that in the analysis of chaotic dynamic systems, dense spatial and temporal sampling can allow the inference of system states “beyond the edge of chaos” predicted from the Lyapunov duration of the system [15], and allow the detection of hidden oscillators and interactors in large complex systems [16]. In a physiological context, Complexity-inspired time series tools have demonstrated utility in detecting the non-linear influence of interventions on organismal physiology. In all these cases, the value of the Complexity approach can be framed as choosing the correct “coarse-graining” of the physiological system (in time and signal-space) such that downstream unsupervised learning approaches can have rich inputs. It is interesting to note that the idea of using multiple densely-spaced points to infer distal system dynamics is found in both Complexity, as well as Traditional Chinese Medicine (e.g. sampling the pulse at multiple points along the wrist).



Works cited, Goal 1 Technical Plan:

1. Spasić, S. Z. & Kesić, S. Editorial: Nonlinearity in Living Systems: Theoretical and Practical Perspectives on Metrics of Physiological Signal Complexity. *Front. Physiol.* **10**, 298 (2019).
2. Mazloom, R. Possibility of Assessing the Gastrointestinal Tract as a Complex System. *J. Med. Syst.* (2019).
3. Wegner, L. H. Modularity Versus Emergence: How to Cope with Complexity in Whole-Plant Physiology? in *Emergence and Modularity in Life Sciences* (eds. Wegner, L. H. & Lüttge, U.) 75–95 (Springer International Publishing, 2019).
4. Robert, C. P. Bayesian Computational Tools. *Annu. Rev. Stat. Appl.* **1**, 153–177 (2014).
5. Friston, K. J., Diedrichsen, J., Holmes, E. & Zeidman, P. Variational representational similarity analysis. *Neuroimage* 115986 (2019).
6. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. Active Inference: A Process Theory. *Neural Comput.* **29**, 1–49 (2017).
7. Satti, R. *et al.* The Application of the Extended Poincaré Plot in the Analysis of Physiological Variabilities. *Front. Physiol.* **10**, 116 (2019).
8. Jelinek, H. F., Cornforth, D. J., Tarvainen, M. P. & Khalaf, K. Investigation of Linear and Nonlinear Properties of a Heartbeat Time Series Using Multiscale Rényi Entropy. *Entropy* **21**, 727 (2019).
9. Kumar, A. *et al.* Development of a heart rate variability and complexity model in predicting the need for life-saving interventions amongst trauma patients. *Burns Trauma* **7**, 12 (2019).
10. Thomas, J. & Zaytseva, A. Mapping complexity/Human knowledge as a complex adaptive system. *Complexity* **21**, 207–234 (2016).
11. Sherwin, W. B., Chao, A., Jost, L. & Smouse, P. E. Information Theory Broadens the Spectrum of Molecular Ecology and Evolution. *Trends Ecol. Evol.* **32**, 948–963 (2017).
12. Wissner-Gross, A. D. & Freer, C. E. Causal entropic forces. *Phys. Rev. Lett.* **110**, 168702 (2013).
13. Smith, B., Kumar, A. & Bittner, T. *Basic Formal Ontology for Bioinformatics*. (IFOMIS Reports, 2005).
14. Brier, S. How Peircean semiotic philosophy connects Western science with Eastern emptiness ontology. *Prog. Biophys. Mol. Biol.* **131**, 377–386 (2017).
15. Alfaro, M., Fuertes, G., Vargas, M., Sepúlveda, J. & Veloso-Poblete, M. Forecast of Chaotic Series in a Horizon Superior to the Inverse of the Maximum Lyapunov Exponent. *Complexity* **2018**, (2018).
16. Shimizu, T. P. & Takeuchi, K. A. Measuring Lyapunov exponents of large chaotic systems with global coupling by time series analysis. *arXiv [nlin.CD]* (2018).

Technical Plan – Goal 2

Personal health tracker devices can provide high-resolution, longitudinal physiological data. Using the generative model discussed in Goal 1, we aim to adapt and improve existing machine learning algorithms to discover multi-sensor personal health predictive markers.

Task 1: Data collection

We will use commercially available non-invasive health sensors on a small number of healthy human subjects. The main sources of information tracked by these mobile devices will include movement (Fitbit, Oura), heart rate (FitBit, Withings BPM core), body temperature (Oura, regular thermometers), blood oxygen level (iHealth air), EEG (Muse 2) and ECG (Withings BPM core), sleep (Fitbit). We will use the Shimmer (https://github.com/openmhealth/shimmer) framework to extract standardized format data, and Muse for the EEG data. We will also be collecting metadata from these subjects including nutrition, exercise and disease states. As opposed to using public data that might be more informative for connecting wearable data to biomarkers and disease states, these datasets will allow us to establish baseline models in a controlled environment.

Task 2: Pattern Discovery

Because we will be collecting from multiple data sources that have different time scales, discovering the structure of these data will require the detection of measurable units. We need to take into account the correlational structure from different data sources through the generative model discussed in Goal 1, as well as the Markov transitions between time points. To reduce the chances of overfitting the physiological model, we will first reduce the time series data into basic measurable units using unsupervised learning. There are certainly a mélange of meaningful time scales present in the data, and Models such as Autoregressive Hidden Markov Chain could be used for the detection of relevant time scales in understanding the data structure. Another technical challenge is the presence of missing data, since some data can be continuously collected (such as heart rate) while others can only be collected during sedentary states (such as EEG) for a comparatively shorter period of time. Methods such as data masking used in time series prediction via Long Short Term Memory (LSTM) learning algorithms may be able to address this problem.

Task 3: Anomaly Detection

To validate that the patterns discovered in Task 3 are representative of the physiological states, we will use anomaly detection to find time points that violate neutral expectations and use the collected metadata as a test set to validate these time points. Anomaly detection is computationally well-established and can be achieved by using algorithms such as isolation forest. The addition of multiple sensor data streams may be able to improve anomaly detection, since different physiological systems will provide mutual information on bodily states.

Key differences in our approach and technical risks

Our approach is different from past approaches that mainly focus on predicting pre-defined states, because we avoid arbitrarily defining observed states that might not have one-to-one mapping to physiological states. In addition, the dataset collected will be multimodal, thus providing a rich interaction and correlational structure that can be very informative for parameterizing the generative physiological model studied in Goal 1. This can potentially allow us to discover much more fine-grained physiological states than current studies. Some of the variables are measured by more than one device in our experimental design, allowing us to cross-validate the data quality from single devices.

The technical risks for a more complex model and multimodal datasets include overfitting, multiple time scales and missing data. Most of these risks are addressed in detail in each task. Another technical risk that pertains to the overall goal is the lack of biomarker measurement, which will be addressed in Goal 3.

Summary

Non-invasive, high resolution and longitudinal physiological data are increasingly easy to capture due to the technological advancements in the consumer wearables markets. Despite a plethora of studies, a model that integrates multimodal physiological data through their biological mechanism is still lacking. We will apply our Complexity Physiology framework (Goal 1) to this task. We will collect multimodal data from multiple decives, . and apply our developed theoretical models of complexity and explore their fit and predictive power, rapidly refining our models and spurring new theoretical insights in a virtuous cycle of discovery. Part of this effort will be developing new tools to improve both descriptive sand statistical models building upon probabilistic programming and machine learning advances. With these refinements to our model and tools, we will be prepared to apply our insights to the high impact areas of human health outlined in Goal 3.

Technical Plan – Goal 3

Human physiology is complex and a loss of this complexity has been proposed to be in part behind human pathologies, including those related to aging (Lipsitz and Goldberger, 1992). Here we extend our work in Goal 2 by going beyond our baseline of healthy physiological complexity and applying our framework of physiological complexity to human diseases. By drawing upon publicly available datasets and training our generative model as a classifier, we will assess the accuracy of our theoretical framework in predicting health outcomes.

Task 1:

We will utilize the large public databases found in Physionet (Goldberger et al., 2000) that contain individuals suffering from cardiovascular disease (e.g. congestive heart failure), chronic neurodegenerative disease (e.g. Parkinson's disease) and healthy control individuals. These datasets provide rich time series data that includes ECG, EEG, interbeat interval, and gait measurements. However, the datasets are collected under diverse settings and different standards. Although we anticipate this diversity adding robustness to our classifier, we also anticipate spending significant time cleaning the data to a common standard, removing missing entries, and compiling the datasets into a common internal database customized for our downstream analysis.

Task 2:

Next will apply the theoretical framework, developed in Goal 1 and encapsulated in our generative model, to these time series datasets. This will necessitate using descriptives and approaches resulting from Goal 2, such as visualization aides and probabilistic programming. This will allow us to quantitatively and qualitatively assess each dataset in the context of our healthy baselines and consider strategies to further process the publicly available data (e.g. batch correction) for feeding into our model. This preliminary analysis will not attempt prediction of diseased states, but rather our model's ability to incorporate the cleaned datasets and return useful descriptives.

Task 3:

The resulting processed internal database and previously developed generative model will then be augmented by a Random Forest classifier to distinguish between individuals belonging to either diseased or healthy control groups given their physiological time series data. The combined model will be trained on a randomly assigned subset of individuals from each dataset, and our combined model's accuracy will be evaluated against the remaining test group within each dataset. We chose the relatively simple Random Forest classifier due to its combination of high accuracy and high explainability, allowing us to inspect the relative importance of emissions produced by our generative model and use those insights to drive further refinement

of our theoretical contribution in Goal 1. As a future direction, we anticipate that with validation of our theoretical framework and this proof-of-concept classifier, we would begin to explore more accurate, albeit less explainable, machine learning approaches that could maximize our impact on human health.

Key differences in our approach and technical risks:

As opposed to a focus upon summary statistics that lose the richness of the underlying longitudinal time series data, we will rely heavily on a expertise-driven generative model as the primary driver for our developed classifier. We will combine this proven approach of joining a generative model and a machine learning classifier with a unique physiological complexity context, allowing us to push into unproven complexity theory while benchmarking our results against more conventional, typically frequentist statistical approaches.

We anticipate a number of technical risks while pursuing our goal. One primary concern is the quality of the publicly available datasets, as extremely noisy data will mask the signal produced by our generative model. The publicly available data we would be using is peer-reviewed Class I and II datasets from the Physionet database and we have allocated sufficient resources in Task 1 to fully clean, apply batch correction, and compile these datasets into our internal database.

Another anticipated risk is that poor performance by our classifier could hinder further refinement of our theoretical framework. We chose the Random Forest classifier for its explainability, but in the event of low accuracy (e.g. less than 60% of individuals correctly classified within the test set), we would explore potentially more accurate but less explainable classifiers such as Support Vector Machines and Deep Neural Networks.

Lipsitz, Lewis A., and Ary L. Goldberger. "Loss of 'complexity' and aging: potential applications of fractals and chaos theory to senescence." Jama 267.13 (1992): 1806-1809.

Goldberger, Ary L., et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." Circulation 101.23 (2000): e215-e220.

Management Plan

Capabilities/Management Plan: The team was formed during ComplexityCon, a convention modeled after the Santa Fe Institute's Summer School. As the conference attracted participants from academia and industry, we are an independent working group and are not affiliated with any single organization. Since then we have been developing our thinking on this topic through research, discussion, and posting on Polyplexus. Our team members have researched the molecular basis of aging and gene expression, worked with enterprise-grade computational resources, and found industry applications for finance and big data technologies. Our PI, Daniel Friedman, recently completed a Ph.D. in Evolution & Ecology at Stanford University, and organized the Stanford Complexity Group. Due to our diverse backgrounds and connections, other experts may be consulted or added to the team as the project progresses.

Meeting Schedule

Our group uses cryptographically-secure coordination software to share chat conversations, files, and software repositories. We will stay synchronized via communication through this channel, as well as all-member video conferences every 2 weeks, in addition to ongoing exchanges through email and version control software.

The majority of our team is located in the San Francisco Bay Area, and we will host an all-hands meeting at least twice during the project. We expect the first meeting to be immediately post-funding, and the second meeting to be following the second ComplexityCon in May 2020. These meetings are included in the Cost Summary, and will be documented in the monthly financial reports. These will contain all expenditures and serve as a token of transparency between CWG and all funding agencies. Additionally, we will create quarterly technical reports documenting our findings, standings, questions, and progress. These reports will be used internally, and made available upon request.

Additionally at the end of the funding term we will have a final review meeting, wherein we describe how we analyze missed/achieved milestones, hold ourselves accountable, discuss any relevant future efforts similar to the award, and plan next collaborations together.

Personnel, Qualifications, and Comments

Andrew McKay
Stanford University
541.980.0620 | amckay1@stanford.edu

Andrew is a PhD student in Biology at Stanford University in the lab of Professor Anne Brunet studying aging. His work focuses on using the African turquoise killifish as a model of vertebrate aging to explore lifespan dynamics and cognitive decline. Prior to his current research, he studied cardiovascular development in the lab of Professor Kristy Red-Horse, received his BS in Biology from the University of Oregon, and his BA in Economics with Honors from Swarthmore College.
Title options: Graduate Student Researcher, Consultant

Colton Cunov
Independent Researcher
214.202.3886 | coltoncunov@gmail.com

Colton Cunov graduated magna cum laude from Texas A&M University with a bachelor's degree in Industrial Engineering.
To investigate cultural influences on crowd dynamics, he worked with faculty at the Indian Institute of Technology Kanpur where he tested how well the most promising physics-inspired models (which were of East Asian and Western origin) performed against data collected from Indian crowds. At A&M he used the character networks of many popular movies and conducted an exploratory analysis on the predictability of the graphs' clique-relaxation values (e.g. k-clique, k-club) on movie ratings from several well-known sources using machine learning methods.
He currently works as a site reliability engineer at a multinational technology managing big data technologies like Hadoop, Kafka, and ElasticSearch.
He expects to spend up to 10 hours per week on this project.

Chenling Xu
University of California, Berkeley.
chenlingantelope@gmail.com | 1-530-220-4220

Chenling Xu was raised in China, started her university career at Zhejiang University, then moved to the United States to study Genetics at the University of California, Davis. After dabbling in bioengineering (International Genetically Engineered Machine competition, 2011, 3rd place) and neuroscience, Chenling started research in evolutionary biology as an undergraduate and continued for a year after graduation with UC Davis Professor Michael Turelli. Chenling worked on projects ranging from the assembly of bacterial genomes, to analyses of genetic variation in plants and humans using statistical models. All these projects

led to publications. Chenling changed her research focus during her PhD at UC Berkeley and developed computational tools for a new molecular technique called single cell RNAseq. The methods she helped develop use techniques such as Bayesian Inference and Variational Autoencoders. This project discovered disease-related markers in single cells using some of these tools. Currently Chenling is in her last year of her PhD, working on a project in collaboration with the Chan Zuckerberg Biohub using single cell RNAseq to discover new cell types and understand aging.

She expects to spend up to 10 hours per week on this project.

Daniel Friedman, PhD.
Independent Researcher & Postdoctoral scholar at the University of California, Davis.
1-650-648-3374 | DanielAriFriedman@gmail.com

Daniel Friedman recently completed his PhD in Ecology & Evolution from Stanford University (finished in August 2019). His dissertation research was under the supervision of Biology Professor Deborah Gordon, and focused on the genetic and neurophysiological basis of the evolution of collective behavior in ants. Specifically the interdisciplinary project involved integrating behavioral, genetic, epigenetic, and neurophysiological datasets to understand general principles of natural selection as it acts on complex adaptive systems. Previous to his thesis research on ant behavioral genetics, Daniel graduated undergraduate with highest honors in Genetics in 2014 from the University of California, Davis. While at UC Davis, Daniel researched the evolution of the genetics of sexual dimorphism in flies with Professor Artyom Kopp, and the regulation of tissue-specific gene expression in honey bees with Professor Brian Johnson.

Also while at Stanford, Daniel organized the Stanford Complexity Group (SCG: complexity.stanford.edu), leading the group from 2016-2019. In SCG, Daniel practiced leadership and memberships in various diverse groupings, organized dozens of events related to Complexity Science, and was exposed to a wide range of colleagues in the field. Daniel attended Santa Fe Institute's Complexity Summer school in 2015, and continues to collaborate with several scientific and entrepreneurial partners he met in Santa Fe. Additionally, Daniel co-organizer the first ComplexityCon in San Francisco in May 2019, where the Complexity Working Group (CWG) was formulated around the current Physiology project.

Apart from his primary research in the evolution, genomics, and neurophysiology of behavior in the social insects, while at Stanford Daniel worked on topics related to philosophy, data analytics, relationships, and art (see publication record/CV).

Currently Daniel is an independent researcher in Davis, CA, and postdoctoral scholar under the mentorship of UC Davis Professor Brian Johnson and U. Pennsylvania Professor Tim Linksvayer. He is most interested in the evolution & regulation of collective behavior in humans and eusocial insects.

He will be a primary worker on the proposed project, dedicating 20 to 30 hours/week.

Richard Cordés
COGSEC
1-631-252-6864 | Richardj.Cordes@gmail.com

Richard J. Cordés formed his first company, The Lexington Investment Group, in 2014 and had been in finance up until that point. After the company took on a neurobiologist as a partner, it pivoted from selling traditional geopolitical and financial research toward systems design and prediction mining and Richard followed suit. Since closing the company to pursue work in these areas, Richard has acted as a Systems Engineering Consultant on projects in a wide variety of fields including the Casino Industry, Finance, Open-Source Intelligence, Threat Analysis, and Education. Richard, during this time, has done research on the optimisation of human learning, memory and neuroscience, data standards, dynamic and emergent ontology, and crowd-engineering. Though he is not a SETA contractor, he currently sits on two committees and chairs a third within Department of Defense projects related to DoD-wide Data Standards and Education Systems Engineering. He is also a member of an IEEE (Institute of Electrical and Electronics Engineers) working group and committee, focused on education and competency data profiles.
He expects to work anywhere between 5 and 20 hours a week on this project.

Capabilities

Our experience is diffused across many fields including biology, ecology, education, and finance. However, we share computational and statistical knowledge that has been applied to our respective areas of concentration. This disparate, yet joint experience gives our team a competitive edge. Our team has no other collective obligations, is not receiving any other funding for this project, and has not proposed this concept to any other government solicitations.

Statement of Work

Period of Performance: The first phase of this project will be carried out from the time of funding, for the following 12 months.

Place(s) of Performance: All tasks will be performed by members of the Complexity Working Group. Deliverables for all phases will include, as noted in the Management Plan, technical and financial progress reports.

Phase I (Months 0-6)

The primary work of Goal 1 (Technical Plan above) is to develop a flexible generative framework for human physiology, drawing on techniques from Complexity Science.

Task 1 – Determine state of the art of Complexity Physiology models.

Milestone 1.1: Comprehensive review of Complexity measures, statistics, and frameworks that could be used for physiological data.

Task 2 – Construct a generative physiology model that can tolerate addition of biological knowledge structures (e.g. dynamic models of physiology, curated bio-databases).

Milestone 1.2: Minimum viable Complexity Physiology model that incorporates dynamical data as well as static knowledge structures.

Task 3 – Demonstrate quantitative improvement of time series classification & prediction through the use of a) generative models as opposed to descriptive models, and b) augmentation of unsupervised methods through biological database.

Milestone 1.3: Provide generative model scaffold that is prepared for input from wearables (e.g. takes in the same data format as publically available device outputs).

Phase II (Month 3-6)

Goals: Collection and cleaning of physiological datasets from healthy subjects, an exploratory analysis on the structure, dynamics, and present time scales of the data, and data cleaning for anomalies via subject metadata.

Task 1 - Collect, extract, and standardize physiological data from healthy subjects

Milestone 2.1 - A preprocessed dataset of continuous physiological data from healthy subjects, lasting many weeks or months

Task 2 - Data analysis and pattern detection

Milestone 2.2 - An understanding of the structure and time scales present within the data and techniques to ensure data integrity.

Task 3 - Anomaly detection and data refinement

Milestone 2.3 - A cleaned, audited dataset incorporating subject metadata to be used in conjunction with public data in Phase III

Phase III (Month 6-12)

Goals: collection and cleaning of publicly available physiological datasets for human diseases, application of developed Theory of Complexity framework-based generative model for unsupervised comparison with healthy baseline, and development of predictive machine learning classifier for control and patient groups.

Task 1 – Process and clean publicly available datasets on ECG, EEG, interbeat interval, and gait for human diseases.

Milestone 3.1 - Internal database of cleaned and standardized human ECG, EEG, interbeat interval, and gait data for human diseases directly comparable to Milestone 2.3 dataset.

Task 2 – Apply developed generative model to cleaned datasets in unsupervised context, evaluating performance using tools built in Goal 2.

Milestone 3.2 - Evaluation of Milestone 1.3 generative model in unsupervised context using Milestone 3.1 internal database of human diseases.

Task 3 – Using generative model connected to Random Forest classifier, create explainable predictor of human disease based upon rich longitudinal time series data.

Milestone 3.3 - Trained and tested classifier built upon Milestone 1.3 generative model with scores for accuracy when applied to internal database from public datasets, broken out by source dataset and disease.

Schedule and Milestones

Estimated schedule:
Goal 1: Start of funding to 9 months post-funding
Goal 2 & 3: 3 months post-funding to 12 months post-funding
Outcomes/Deliverables:
Peer-reviewed publications detailing the advances achieved within each Goal (e.g. one Theoretical paper, one paper with Computational/Statistical/Wearable information, one paper with empirical proof-of concept using the framework and tools described).

Cost Summary

	PHASE 1			PHASE 2		
DIRECT LABOR			**Amount**			**Amount**
1 Full time equivalent (Split 5 ways)			$32,500.00			$32,500.00
*No Subcontractors						
*No Fringe Benefits						
SUBTOTAL			$32,500.00			$32,500.00
MATERIALS & EQUIPMENT	*Price*	*Quantity*	**Amount**	*Price*	*Quantity*	**Amount**
Muse 2 EEG				$250.00	5	$1,250.00
Withings BPM Core				$100.00	2	$200.00
iHealth Air				$75.00	2	$150.00
Oura Ring				$300.00	5	$1,500.00
Fitbit				$100.00	5	$500.00
SUBTOTAL			$-			$3,600.00
OTHER DIRECT COSTS (ODC)	*$ / Period*	*est. Time*	**Amount**			
Storage 2tb SSD/Snapshot backup	$300.00	10 months	$3,000.00			
GPU Computing, p3 2xL AWS instance	$3.06	~2000 hours	$6,000.00			
General Computing, AWS t2.xlarge	$0.19	~5000 hours	$950.00			
Elastic IP 0.01/GB	n/a	n/a	$200.00			
SUBTOTAL			$10,150.00			$-

CONFERENCE TRAVEL, REG., & PUBLISHING			**Amount**			**Amount**
HAPs 2020 Annual Conference						$9,000.00
Travel for in-person Meetings						$3,000.00
SUBTOTAL			$-90,750.00			$12,000.00