Research and Scholarship

My research interests in learning and decision-making have led to a productive thesis on the cognitive costs of behavior in health and disease. I have published 2 first-authored and 3 co-authored papers in reputable journals, built several international collaborations, and successfully funded my own Ph.D. with 3 grants. I have also involved undergraduate and high school students in my research projects as an extension of my teaching endeavors.

Policy Compression in Action Selection

The human brain has limited cognitive resources for use in everyday behaviors such as learning and decision-making. Behaviors can be cognitively costly because of the amount of mental effort that is required to perform a task. Cognitive cost impacts performance in many domains, from low-level functions like visual memory all the way up to high-level cognitive tasks, such as studying for an exam. **How do humans weigh the cognitive costs of behavior against its potential benefits or rewards? And how does this trade-off between cognitive effort and reward impact behavior in health and disease?**

To address these questions, my research utilizes behavioral experiments and computational modeling. One challenge in studying this cost-benefit trade-off is that mental effort can be characterized in a variety of ways: computational models have formalized effort as the amount of thinking time that is required for a task, as well as the inherent difficulty of learning a task (Dayan & Daw, 2008; Filipowicz et al., 2020). But less studied is the **amount of *memory* required for executing a task**. My research aims to fill this gap by examining how humans weigh rewards against the memory demands imposed by behavior.

First, I developed a computational model describing how people should act if they seek to maximize rewards under a fixed memory capacity. This model uses information theory to quantify memory load as the amount of environmental “state” information that is taken into account when selecting actions. Under this framework, optimal behavior achieves a balance between the complexity of the behavioral policy, or how specific one’s actions are to the states, and the amount of reward that can be obtained (Parush et al., 2011; Still, 2009; Tishby & Polani, 2011). When behavior is simplified by reducing demand on memory, it is called **policy compression**.

Next, I explored the consequences of policy compression in a variety of tasks. I showed that a range of behavioral phenomena including habit formation, action chunking, and navigational planning could be described in terms of reward optimization under limited memory capacity. Additionally, our model predicts that maladaptive behaviors such as perseveration, which is implicated in various psychiatric diseases, are a consequence of reduced memory capacity. We validated these model predictions in a previously published dataset, showing that patients with schizophrenia are biased towards less complex policies compared to healthy controls (Gershman & Lai, 2021).

In novel behavioral tasks, I test further predictions of our model to reveal how people adapt to varying incentives and demands on memory. I showed that action chunking, or the act of grouping individual actions together to enable faster execution, is modulated by cognitive load and can be understood as policy compression (Lai et al., 2022). In collaboration with the Janak Lab at Johns Hopkins, I am applying our policy compression framework towards understanding habit formation under different reward schedules.

The outcome of my thesis research has important applications to the field and to society: knowing how humans specifically trade-off memory load and reward is important for developing a more holistic model of resource-rationality in human behavior. My work could inform the development of interventions for psychiatric

conditions associated with memory deficits, and could also inspire the design of human-centered technologies that alleviate memory load for easier decision-making.

A Computational Account of Egodystonia

During my PhD, I developed an interest in computational psychiatry and started a collaboration with Quentin Huys and Tobias Hauser at the Max Planck Institute for Computational Psychiatry and Ageing in London. Our project focused on a curious behavioral phenomenon called egodystonia: a metacognitive problem where one's actions and their subjective accounts are detached from one another (Robbins et al., 2019). Egodystonia has been documented in psychiatric conditions such as obsessive compulsive disorder (OCD), bulimia nervosa, and drug addiction.

What drives this mismatch between beliefs and behavior? Neuroscientists have shown that people with OCD are able to develop accurate beliefs about the environment but fail to use their beliefs to guide their actions (Vaghi et al., 2017, 2019). While these studies have shown a dissociation between cognitive knowledge and subsequent behavior, none have (1) provided an explanation for how this mismatch arises or (2) measured one's subjective sense of their actions being "bothersome," which is a key signature of egodystonia. We wondered if we could elicit egodystonic feelings in a healthy population as a key to understanding why and how it emerges.

In a novel experiment combining behavior and subjective report, we successfully elicited egodystonic feelings in a healthy population with a range of obsessive compulsive traits. We found that egodystonicity does not change with reward availability or action rate, and is driven by a fear of consequences or "missing out." Guided by our results, we develop a computational account of egodystonia that can capture our behavioral and subjective report data. Our results provide the first evidence for experimentally-induced egodystonia and paves the way for a better understanding of this curious phenomenon.

Future Research Agenda

Going forward, I hope to expand the scope of my research towards understanding how cognitive limitations can lead to severe behavioral consequences such as belief rigidity and polarization. This is an important and relevant subject given the recent polarization of political ideologies in the United States, as well as the effects of individual beliefs on climate (in)action. It is also pertinent for understanding the mechanisms underlying disordered beliefs and actions in psychiatric disease. Understanding the influence of cognitive resource limitations on belief formation and maintenance is therefore crucial for developing interventions to change beliefs.

I believe that the UCSD is a great place to carry out these research interests. In collaboration with talented undergraduates and research faculty with similar interests (such as Marcelo Mattar, Ed Vul, Adam Aron, and Craig McKenzie), I will continue to pursue my curiosities about brain and behavior as both a researcher and an educator. I also hope to help train and inspire the next generation of scientists by bringing star undergraduates to conferences and sitting on graduate student thesis committees. Finally, I will continue to share my research and scholarship with the broader community by volunteering to give talks at local high schools (as I have in the past) and writing popular science articles.

References

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. Cognitive, Affective & Behavioral Neuroscience, 8(4), 429–453.

Filipowicz, A. L. S., Levine, J., Piasini, E., Tavoni, G., Kable, J. W., & Gold, J. I. (2020). The complexity of model-free and model-based learning strategies. In bioRxiv (p. 2019.12.28.879965). https://doi.org/10.1101/2019.12.28.879965

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8410–8415.

Gershman, S. J., & Lai, L. (2021). The Reward-Complexity Trade-off in Schizophrenia. Computational Psychiatry, 5(1), 38–53.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. Theory into Practice, 41(4), 212–218.

Lai, L., Huang, A. Z., & Gershman, S. J. (2022). Action chunking as policy compression. https://doi.org/10.31234/osf.io/z8yrv

O'Brien, J. G., Millis, B. J., & Cohen, M. W. (2009). The Course Syllabus: A Learning-Centered Approach. John Wiley & Sons.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. Educational Psychologist, 38(1), 1–4.

Parush, N., Tishby, N., & Bergman, H. (2011). Dopaminergic Balance between Reward Maximization and Policy Complexity. Frontiers in Systems Neuroscience, 5, 22.

Robbins, T. W., Vaghi, M. M., & Banca, P. (2019). Obsessive-Compulsive Disorder: Puzzles and Prospects. Neuron, 102(1), 27–47.

Still, S. (2009). Information-theoretic approach to interactive learning. EPL, 85(2), 28005.

Tishby, N., & Polani, D. (2011). Information Theory of Decisions and Actions. In Perception-Action Cycle (pp. 601–636). https://doi.org/10.1007/978-1-4419-1452-1_19

Vaghi, M. M., Cardinal, R. N., Apergis-Schoute, A. M., Fineberg, N. A., Sule, A., & Robbins, T. W. (2019). Action-Outcome Knowledge Dissociates From Behavior in Obsessive-Compulsive Disorder Following Contingency Degradation. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 4(2), 200–209.

Vaghi, M. M., Luyckx, F., Sule, A., Fineberg, N. A., Robbins, T. W., & De Martino, B. (2017). Compulsivity Reveals a Novel Dissociation between Action and Confidence. Neuron, 96(2), 348–354.e4.

Wiggins, G., Wiggins, G. P., & McTighe, J. (2005). Understanding by Design. ASCD.