How does the mind connect with the physical world? This question constitutes one of the most tantalizing remaining frontiers of human knowledge. Both scientific and philosophical, it may never be fully solved, but valuable progress has refined our understanding of the problem, contributed a wealth of empirical data, and provided new approaches and theories. My research makes concrete contributions to this progress by developing embodied cognitive computer models. In addition to the grand intellectual challenge, this research is motivated by application areas ranging from medicine (e.g., rehabilitation and treatment of cognitive deficits in humans) to robotics (e.g., assisted living and autonomous disaster recovery).

My research makes concrete contributions to this progress by developing embodied cognitive computer models. In addition to the grand intellectual challenge, this research is motivated by application areas ranging from medicine (e.g., rehabilitation and treatment of cognitive deficits in humans) to robotics (e.g., assisted living and autonomous disaster recovery).

My vision for this research involves several specific elements. First, an emphasis on physical embodiment, which is crucial for understanding how cognition is grounded in sensory perception and impacts motor control. Second, the development of mathematical models with explanatory power that go beyond statistical correlations in explaining how neural computations give rise to cognition. In particular, the long-term focus is neurocomputational representations of structured data and reasoning, using inference strategies that are learned from experience rather than encoded by human domain experts. As such, my research is necessarily collaborative and interdisciplinary, drawing on knowledge from computer science, robotics, cognitive science, neuroscience, and mathematics. The following two sections describe my research contributions as a doctoral student, which begin to independently address certain elements of this long-term research vision. The last section describes how I will continue to build upon my existing research in my current postdoctoral and future work.

Robotic Imitation Learning via Cause-Effect Reasoning

My dissertation focused on causal reasoning models that support cognitive-level robotic imitation learning (IL). Robotic IL enables human end-users without computer programming ability to show a robot how to behave, by demonstrating the behavior themselves. While sensorimotor IL is concerned with the teacher’s precise motions, gestures, and dynamics, cognitive IL is concerned with the teacher’s intentions, independent of the requisite motor actions in any given situation. This enables generalization to new situations where the imitator must operate within a different environment and embodiment than the teacher.

In my doctoral work [2–5], I modeled intention inference as a causal reasoning problem, wherein hidden intentions are viewed as causing observable actions. I developed novel causal inference algorithms with formal correctness and complexity guarantees that were capable of solving this problem, by leveraging background knowledge supplied by a domain expert. Empirical experiments showed that a physical robot using this framework, as well as human learners, could both generalize learned skills on the basis of a single demonstration. These findings, along with existing cognitive and neuroscientific evidence, suggest that this framework may be a reasonable model of certain cognitive processes in humans.

This work fits with my long-term research goals because it has honed my expertise with embodied systems and establishes a reasoning model that is relevant to artificial and potentially biological cognition. On the other hand, it is based on symbolic (non-neural)
computation, and relies on background knowledge that is encoded by a human. My future work will pose this model as a target for a neurocomputational learning system.

**Locating Fixed Points using Directional Fibers**

In parallel with the foregoing, I developed a mathematical tool to help study recurrent neural network dynamics [1]. This tool is a first step towards a neurocomputational implementation of the IL framework described above. In particular, I defined novel mathematical objects called “directional fibers,” which comprise a family of curves within the state space of a dynamical system such as a recurrent neural network. I showed that directional fibers will contain many fixed points of their respective dynamical system, and that under commonly satisfied conditions, they can be numerically traversed to systematically enumerate those fixed points. These theoretical results were borne out with empirical computer experiments.

This work is significant because it provides a new way to systematically locate many fixed points in many dynamical systems. Fixed points are some of the most fundamental features of a dynamical system, but they can also be viewed as solutions to non-linear equations or zeros of a gradient field, so directional fibers are also more generally applicable to non-linear equations and optimization. In terms of my long-term research goals, this fixed point solver provides a new perspective on and a useful tool for studying recurrent neural dynamics, which can potentially lead to new theories of the neural basis of cognition. However, the current state of this work is still rather abstract, and its full potential depends on some open mathematical questions. My future work will investigate these open questions further, and apply this method to concrete examples of embodied neurocognitive models.

**Postdoctoral and Future Research**

My current postdoctoral research focuses on cognitive disorders in humans. The primary objective is to build a neurocomputational model that captures key cognitive deficits involved in post-traumatic stress disorder (PTSD). This model will be validated against existing data, used to investigate existing hypotheses about the neurological basis of PTSD, and used to generate new testable predictions that can be studied clinically. Testing and development of the model will use cognitive tasks commonly administered to PTSD patients, which involve executive control over attention, working memory, and goal-directed reasoning.

To implement models like this in my postdoctoral and longer-term research, I plan to further develop and combine my existing work with recent techniques for “programmable” neural networks developed by other researchers (e.g., [6, 7]). These techniques begin to enable neural networks to execute symbolic algorithms more commonly implemented using traditional computer programs. However, they tend to rely on non-neural components, problem-specific architectures, human-authored knowledge, and/or learning algorithms with limited interpretability. Moreover, none of this past work addresses physical embodiment. There is a need for a single embodied model that is purely neurocomputational, can store multiple arbitrary algorithms, can learn from experience, and is based on intelligible, well-understood operating principles. I plan to develop such a model by combining the strengths of other researcher’s programmable neural models with my own work on embodied cognitive systems and mathematical analysis of neural dynamics. Over the long term, my hope is that models like this will ultimately improve our understanding and engineering of neural computation and embodied cognition.
References


Acceptance rate: 19%


Best Student Paper Award
