Research Statement

Will Bridewell

Research on *computational scientific discovery* has enriched the field of artificial intelligence for over four decades, inspiring early work on expert systems, machine learning, and data mining. Nevertheless, this area remains distinct due to its grounding in established scientific domains and their native representations. This emphasis leads to research that ranges from prescriptive scientific methodology to practical, goal-directed applications. My own work centers on technology for modeling complex, dynamic systems.

Whether they use mathematical equations, interacting agents, or other formalisms, scientists build models to explain empirical findings and to predict future behavior. Routinely these models execute on computers to calculate species interactions, chemical kinetics, climate change, and more. Analyses of the resulting simulations increase our understanding of complex systems and may ultimately inform public policy. Consider an ecologist studying the biological environment of a local estuary. In the course of her work, she gathers data from satellite images, observational stations, and controlled experiments with each source yielding a piece of the puzzle. To see the whole picture, she models environmental variables (e.g., surface runoff, plankton concentration) and their relationships (e.g., growth limitation, predation) and uses that characterization to discover and communicate knowledge. Although models such as this can deepen our understanding of the world, model building is a tedious and error-prone process that requires substantial background in mathematics, programming, or both. These barriers limit one's ability to explore plausible alternatives and consume resources better spent on discovery and analysis than on implementation details.

My research on *inductive process modeling* addresses these problems through systems that automate model construction and revision. Given time series and domain-specific knowledge, such systems search the dual space of candidate structures and their numerical parameters to discover the most plausible explanations of the data. The underlying formalism, which combines high-level conceptual knowledge with lower-level mathematical expressions, lets scientists easily draw from libraries of predefined components or create their own. This approach integrates research from multiple areas of artificial intelligence and scientific computing to open new avenues for the principled consideration of alternative models. Moreover, the process-based representation fits well with biology, chemistry, and other sciences.

A central problem in inductive process modeling involves implausible explanations. Take for example a model of predator–prey dynamics that lacks a means of predation. The idea itself is absurd, but without constraints to limit its search, a modeling system would consider that case. To address this problem, my colleagues and I have developed formalisms for expressing these constraints and an approach that learns them. Encouraging results show that the inferred constraints not only reduce search in similar modeling scenarios but also transfer across domains and correspond to scientific knowledge.

Moving forward, my sights are focused on several exciting challenges. For instance, the ability to consider thousands of alternative explanations and identify commonalities among the promising ones opens the door to a new computational methodology of science. To bring this technology to a wider audience, we must expand modeling systems to work with agent-based formalisms, cellular automata, and other common scientific representations. In addition the power of this method may help build integrated models that cut across levels of abstraction—for instance, linking plant-level dynamics to local climate change. Such model integration brings with it corresponding advances in data integration. Rounding out this picture, I see the tight coupling of modeling and constraint induction leading to systems that adapt to their tasks.

Although computational scientific discovery has received little attention, it naturally bridges the fields of eScience and intelligent systems to transform computers into creative partners in research. This broad vision depends on systems that increase their utility and value by adjusting to researchers' needs. Ultimately, I look forward to a time when software environments routinely collaborate with scientists to forge new knowledge.