Project Summary

This project will develop a unified theory of how natural and artificial systems can learn to solve complex motor tasks, such as running, diving, throwing, and flying, that entail significant sensory input and the coordination, sequencing, and fine-tuning of many low-level activities. Such a project is possible because of significant experimental advances in our understanding of motor control systems in humans and other animals, and because of increased sophistication in our mathematical models of control learning. These models will be used not only to analyze and predict natural phenomena in motor control, but also to derive effective adaptive controllers for artificial systems carrying out complex tasks. In particular, a microelectromechanical winged insect will be trained to fly.

To generate complex behaviors, natural and artificial systems must be organized hierarchically with multiple layers of abstraction. The first research task will therefore be to identify appropriate levels of representation at which the physical system can be modelled and at which control actions can be defined. For example, in describing an insect flying from A to B, possible levels might be 1) nerve signals and mechanical properties controlling the detailed shaping of each wingbeat 2) basic wingbeat cycle 3) "steering" the cycle to direct flight 4) takeoff, navigation, landing. Detailed motion, force, and airflow measurements will be made under a variety of experimental circumstances and flying tasks to establish the correspondence. Similar experiments will be carried out for running in cockroaches and for running, diving, and throwing in humans. These studies (and, in the case of insects, neurophysiological studies) will also establish the sensory inputs that are available at each level of the control system.

Given the general structure of the control system and the appropriate sensory inputs, the next step is to design learning algorithms capable of learning to perform the given task successfully. The learning method to be used is *reinforcement learning*, a technique designed to adjust the control algorithm to optimize an *objective function*–that is, the long-term accumulated value of a specified *reward signal*. The reward is supplied to the learning algorithm as part of the sensory input. New reinforcement learning algoriths will be developed that operate using both local and global reward signals within a hierarchical control structure; furthermore, these algorithms will be proved to converge even using nonlinear representations of the overall objective function.

It is hypothesized that natural control systems may be implicitly optimizing an objective function, and that some systems may use a form of reinforcement learning from a reward signal to do so. It is not known what the reward signal might be, but the failure of previous simpler proposals suggests that it must be a complex, nonlinear, multiattribute combination of the sensory inputs. The final step in establishing correspondence between natural and artificial systems is therefore to ascertain what it is that the natural control system is implicitly optimizing, if anything—that is, to solve the *inverse reinforcement learning* problem of determining the reward function by fitting the mathematical model to observed behavior in a variety of circumstances. This task is a new form of computational learning that has been little studied outside one specialized subdiscipline of econometrics, but which is of broad significance in modelling purposive behavior of all kinds.

The team will derive algorithms for inverse reinforcement learning, and will show mathematically the circumstances under which the task is soluble. They will then carry out an experimental program for the motor learning tasks and natural systems described above, including the use of ascertained reward functions to predict system response under new circumstances. This research should shed light on the central question of whether this form of learning in animals and humans can be viewed as driven by optimization or by some other principle, such as the preservation of fixed interface characteristics among the various levels of the system. Discovery of consistent reward functions in animals, especially humans, would have significant consequences for general theories of learning.