Technical Summary of Proposed Research
Lynn Andrea Stein

Architectures for Cognitive Robotics

My area of research is cognitive robotics—the attempt to construct a "smart" system based on the interaction between body and "mind". In contrast to traditional robotics, which focuses on the engineering of the "body" and low level control systems, cognitive robotics aims at the kinds of intelligent behavior that we are inclined to call "thinking". To accomplish this, the field draws on both engineering practice and scientific knowledge of extant intelligent systems: human – and, to a lesser extent, animal—biology, neurophysiology, and psychology.

This area of research has been active for approximately the last decade, though its roots are far deeper (including Norbert Wiener's Cybernetics and much of the philosophical tradition). In its current manifestation, it is largely though not entirely an outgrowth of artificial intelligence, itself a subfield of computer science. Its successes have been limited but not insignificant. There are now robots that interact successfully with humans in limited contexts or at limited skill levels (e.g., navigation and tour-giving). However, many of these systems are viewed as "one-ofs", engineering solutions with little learned that is directly applicable to the next problem. In particular, there are serious questions as to how scalable successful robotics techniques are into the realm of cognition.

What is needed now is principled argument as to how these next-generation cognitive robotics systems can be constructed. Unfortunately, the principles most often espoused for these problems are drawn from traditional computer science and most if not all rely on the idea of *modular-functionalism*: that structural and functional decomposition are one and the same. Experience has given us reason to question the applicability of modular-functionalism to cognitive robotics. In my project, I propose to explore an alternative approach to architectural principles for cognitive robotics, one based on *neo-modularity* rather than modular-functionalism. In particular, I will construct a system designed according to these principles and use this system as well as its many progenitors as case studies in a book and other writings which lay out the principles of neo-modularity as an alternative to the more traditional modular-functionalism.

In the remainder of this proposal, I first outline computer science's idea of modular-functionalism, contrast it with other disciplines, and describe why modular-functionalism may be inapplicable to cognitive robotics. Next, I present alternatives, in the form of neo-modular principles, which I believe form a more suitable foundation for cognitive robotics. Finally, I describe the system that I propose to build while at the Bunting Institute. More technical details of my agenda can be found in the accompanying paper, "Neo-Modular Systems: Architectural Principles for Cognitive Robotics."

Computer Science and Modular-Functionalism

To understand the ways in which my work is a departure from traditional computer science, it is necessary first to understand the ideas of structural and functional decomposition. Structural decomposition refers to the natural decomposition into "pieces" or components of the system. For example, a cup might logically decompose into its handle and its bowl. Alternately, we can talk about the functional decomposition of the cup: it can be used to hold liquid, and it can be carried. In this case, there is a clear coupling between the structural and functional decompositions of the system: the handle admits the carrying function, while the bowl is responsible for the holding-liquid part of the job. If the bowl were to break, the handle would still work, and vice versa.

Contrast this with a simple wind-up toy which periodically goes around in circles. The way that these are typically made is with a winding mechanism, two wheels, and a lever on one of the wheels. The lever rotates with the wheels; periodically, it comes into contact with the ground and lifts its wheel from the ground, causing the toy to turn. In this case, the wheels (together with the winding mechanism) are responsible for going straight; but the turning behavior is a property of the interaction between the lever and the wheels. If the lever were to break, the toy might still go straight, but the lever itself is not sufficient to turn the toy.

This second example illustrates a coupled system. The structural components of the system do not correspond directly to its functional behavior. This is a fairly simple example, but many more complex ones exist. Most of engineering deals with coupled systems of this sort, as does modern biology, e.g., in the common dissociation between anatomy (body structures) and physiology (body functions).

Computer science does not.

Almost all of computer science is based on the idea that there is, and ought to be, a strong correspondence between functional and structural decompositions. This idea is typically taught in introductory computer science courses as well as software engineering and industry-standard design. There are many good reasons to use this kind of decomposition in artificial, disembodied computational systems, and the ideas have proved valuable in these domains. [1]

But cognitive robotics is different.

In cognitive robotics, much of the power comes from layering behaviors one on top of the other in much the way that the wind-up toy's lever modifies the behavior of the wheels without imposing its own separately viable functionality. Brooks's Subsumption Architecture was an attempt to argue this point broadly; recent interest in dynamical systems (e.g., Beer) is another example. These approaches have proved valuable in designing and constructing increasingly sophisticated "bodies". The challenge now is to significantly advance the level at which these systems operate, to give them apparent cognitive capacity.

Unfortunately, much work along these lines has taken the approach to be expected of computer scientists: layering separate "thinking" systems on top of these robotic bodies. Since the two typically need a separate component to mediate between them, the resulting (prevailing) approach to cognitive robotics has come to be known as a *Three Level Architecture*, or *TLA*: a "thinking box" wired onto the top of a robotic "body". Although it is built out on two reasonably successful fields of research—robotics and traditional artificial intelligence—this approach to cognitive robotics has been severely limited. I would argue that this is because it is based on modular-functional design principles, which really aren't all that appropriate to the cognitive robotics endeavor.

An Alternative Architectural Approach: Neo-Modularity

Since my work is concerned more with the cognitive ends of the system, I have spent much of the last several years studying existing systems that function at the robotic/reactive level (e.g., don't bump into things) and extending some of these to more "intelligent" behavior *without* building TLAs. For example, one of my projects involved taking a navigation robot that could return to previously visited locations and enabling it to use a map to get to a novel location; another involved taking a navigation robot and allowing its human companion to "show it around" the office, after which the robot could navigate the office environment itself even though many of the landmarks—such as "the poster of Mars on the wall"—were invisible to the robot's sensors.

[1] Footnote: Interestingly, as recent computational systems have moved towards networks and interactive systems, this claim may be weakening even in its traditional domains. I and others have been advocating an introductory computer science course which bridges these two approaches, teaching both modular-functional and coupled/interactive techniques.

Both of these projects involved building on the base robot ("body") behavior, directly exploiting skills that were necessary to navigate to perform more cognitive tasks. In the process of this research, I came to the realization that there are indeed significant principles of cognitive robotics which can be generalized from one project to the next. Indeed, these principles are empowering both for design and for analysis of cognitive robotic systems, and they are based on fundamental aspects of the cognitive robotics problem.

For example, the "body" of the robot—its physical subsystems and its lowest level behaviors, e.g., not bumping into things – necessarily embody understanding about the nature of space. It is essentially impossible to build a navigation robot without giving it some implicit representation of the connectivity of space, if only in terms of permissible actions at any point. But this means that we can leverage that representation to solve other problems as well. This is essentially the insight behind the map-reading system described above: in order to get "back" to a place, the robot had to "understand", in some very loose sense, how space interconnects. To get this system to read maps, I had it pretend that the map was an environment it could explore and build up this "understanding" of a space it had never seen. In other words, by short-circuiting the robot's connections to the world (turning off its sensors and motors), I was able to get it to "imagine" interacting with the map and to use representational ability implicit in its "body" to solve a more abstract (cognitive) problem.

The idea of using the "body knowledge" of a robot to enable cognition has support in psychology, linguistics, and neuroscience. For example, Lakoff and Johnson have argued that language exemplifies the mapping of body-based metaphor onto cognitive domains. Kosslyn makes a strong case that visual imagery implicates precisely the same brain processes (and regions!) as actual vision. And A. Damasio argues that the lower-level "instinctive" portions of the brain are crucial for what is generally regarded as the "highest" level of cognitive capability, rational decision-making.

It's also fairly natural when you think about the opportunistic nature of development. After all, both evolution and individual development rely on systems that are already there in building more advanced systems. While this is not an argument that we *couldn't* in principle rationally redesign evolved systems, it makes a fairly compelling case that we shouldn't expect to see modular-functional systems when we look at natural examples of intelligence. And, since these are the only intelligent systems that we have as models, systems based on the design of natural intelligence aren't necessarily going to follow modular-functionalism either.

Although I don't believe that cognitive robotics is modular-functional, this doesn't mean that it can't have powerful constructive and analytic principles. My recent work has focused on identifying just such principles. In particular, I have culled three from my own work and that of others, both in cognitive robotics and outside the field:

1. The body's interactive capacity to solve a concrete problem (e.g., navigation) can often be leveraged to solve a cognitive correlate (e.g., map following).

2. The world provides a common set of "ground points" which can provide a bridge between different representational systems (e.g., the robot's and the human's).

3. Gradual growth/scaling means that learning and adaptation can be used to move from one solution to another (e.g., from waving an arm at the shoulder to pointing using multiple joints).

Since each of these principles talks about ways in which to build a more complex system out of simpler ones, they are all in some sense "modular". However, none are modular-functional in the sense of computer science: they get their power from the interaction of multiple systems, systems with environments, or systems over time. For this reason, I call these principles *neo-modular*.

Building a Neo-Modular System

My work in the past several years has illustrated these neo-modular principles in a number of systems; others, both within and outside of the field have used them implicitly in their own work. However, no system has so far been built which is designed explicitly using these principles. Such a proof-of-concept example is crucial to the strength of these arguments, since it is only by their empowering ability that such architectural principles can be judged. My project during the fellowship year is to build such a system and use it, along with other existing examples, as case studies in a book and other writings describing this neo-modular approach to cognitive robotics.

I have elsewhere proposed a system based on the combined power of these principles. (See the accompanying paper, "Neo-Modular Systems: Architectural Principles for Cognitive Robotics.") This system explicitly uses these principles to leverage the body, the world, and development to scale from primitive motor-coordination behaviors to plan sequencing to social interaction. At each stage, it uses neo-modular principles in a variety of ways to leverage existing abilities to produce more complex and sophisticated behavior without adding separate, functionally independent components.

The system begins with two basic interactive functions: moving the eye to center a visual target, and moving a simple arm to a fixed point on the periphery of the workspace. It layers a series of developmental levels on top of these:

1. Hand-eye coordination: identify an interesting point visually, then move the hand to it.

2. Complex reaching: move around obstacles and across the workspace. Since this may involve sequencing several motions, it requires some rudimentary planning.

3. Shared attention: when interacting with a person, look at the object that the person is pointing to. This requires understanding the triangle between self, other, and object, as well as enough of the social context to look at the object instead of the other person.

Since the system involves scaling the proportions of the body, it will necessarily be built in simulation: building a robot which physically grows is significantly beyond the capabilities of today's robotics laboratories. Since the implementation work will be done entirely in simulation, it requires only the computer that I will bring with me to the Bunting; no laboratory facilities are needed. (Of course, I have a laboratory at MIT should the need arise; my goal is to avoid it as much as possible during the fellowship year!)

Over the remainder of this academic year, I hope to finalize details of the design of this system and begin some preliminary experiments into its implementation. Thus, I would expect to begin the fellowship year with a computational framework in place, ready to complete the details of the implementation, and eager to make the complete case for these architectural arguments. I expect to have drafted an outline for the book—certain pieces exist in my prior publications and informal sketches—by the summer of 1997, and would plan to write it over the fellowship year.

I believe that the field of cognitive robotics is currently handicapped by its roots in computer science and the discipline's view that "good" design is modular-functional. There are empowering principles for cognitive robotics outside of the world of modular-functionalism; the case needs to be made for these principles and this alternative world view. There are many *post hoc* examples of their power. What remains is to build one from scratch, with these principles at its heart, and to use this example as the centerpiece of a case for neo-modular architectures.

References

[Bates] Bates, E. 1979. The Emergence of Symbols. New York: Academic Press.

[Beer] Beer, R.D. 1990. Intelligence as Adaptive Behavior. Cambridge, Massachusetts: Academic Press.

[Brooks] Brooks, R.A. 1986. A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation 2(1):14–23.

[Crick and Jones] Crick, F. and Jones, E. 1993. Backwardness of human neuroanatomy. Nature 361:109–110.

[Damasio] Damasio, A.R. 1994. Descartes' Error: Emotion, Reason, and the Human Brain. New York: G.P. Putnam's Sons.

[Damasio and Damasio] Damasio, H. and Damasio, A.R. 1989. Lesion Analysis in Neuropsychology. New York: Oxford University Press.

[Dennett] Dennett, D.C. 1991. Consciousness Explained. Boston: Little, Brown.

[Dennett and Kinsbourne] Dennett, D.C. and Kinsbourne, M. 1992. Time and the Observer: The Where and When of Consciousness in the Brain. Brain and Behavioral Sciences 15:183–247.

[Diamond] Diamond, A. 1990. Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of Inhibitory Control in Reaching. In The Development ant Neural Bases of Higher Cognitive Functions. New York: The New York Academy of Sciences.

[Fodor] Fodor, J. A. 1983. The Modularity of Mind. Cambridge, Massachusetts: MIT Press.

[Haugeland] Haugeland, J. 1985. Artificial Intelligence: The Very Idea. Cambridge, Massachusetts: MIT Press.

[Harvey et al.] Harvey, I.; Husbands, P.; and Cliff, D. 1994. Seeing the Light: Artificial Evolution, Real Vision. In Proceedings of the Third International Conference on Simulation of Adaptive Behavior. Brighton, England.

[Hobbs and Moore] Hobbs, J. and Moore, R., eds. 1985. Formal Theories of the Commonsense World. Norwood, New Jersey: Ablex Publishing Co.

[Kinsbourne] Kinsbourne, M. 1988. Integrated field theory of consciousness. In A. Marcel and E. Bisiach, eds., The Concept of Consciousness in Contemporary Science. London, England: Oxford University Press.

[Kosslyn] Kosslyn, S.M. 1994. Image and Brain: The Resolution of the Imagery Debate. Cambridge, Massachusetts: The MIT Press.

[Knight and Luk] Knight, K; and Luk, S.K. 1994, Building a Large-Scale Knowledge Base for Machine Translation. In Proceedings of the Twelfth National Conference on Artificial Intelligence. Seattle, Washington.

[Lakoff] Lakoff, G. 1987. Women, Fire, and Dangerous Things. Chicago, Illinois: University of Chicago Press.

[Lakoff and Johnson] Lakoff, G., and Johnson, M. 1980. Metaphors We Live By.

[Lempert and Kinsbourne] Lempert, H. and Kinsbourne, M. 1985, 'Possible origin of speech in selective orienting', *Psychological Bulletin* **97**, 62–73.

[Lisberger] Lisberger, S. G. 1988, 'The neural basis for motor learning in the vestibulo-ocular reflex in monkeys', *Trends in Neuroscience* **11**, 147–152.

[Littman et al.] Littman, M.L.; Cassandra, A.R.; and Kaelbling, L.P. 1995. Learning Policies for Partially Observable Environments: Scaling Up. In Proceedings of the Twelfth International Conference on Machine Learning.

[Marjanović et al.] Marjanović, M.; Scassellati, B.; and Williamson, M. 1996. Self–taught Visually–guided Pointing for a Humanoid Robot. In Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior. Cape Cod, Massachusetts.

[Matarić] Matarić, M. 1992. Integration of Representation Into Goal-Driven Behavior-Based Robots. IEEE Transactions on Robotics and Automation 8(3).

[Minsky] Minsky, M. 1986. The Society of Mind. New York: Simon and Schuster.

[Mussa-Ivaldi et al.] Mussa-Ivaldi, F.A.; Giszter, S.F.; and Bizzi, E. 1994. Linear Combinations of Primitives in Vertebrate Motor Control. Proceedings of thte National Academy of Sciences 91:7534–7538.

[Perkowitz and Etzioni] Perkowitz, M.; and Etzioni, O. 1995. Category Translation: Learning to Understand Information on the Internet. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. Montreal, Quebec, Canada.

[Ringrose] Ringrose, R. 1994. Automatically Tuning Control Systems for Simulated Legged Robots. In Proceedings of the Twelfth National Conference on Artificial Intelligence. Seattle, Washington.

[Sims] Sims, K. 1994. Evolving 3D Morphology and Behavior by Competition. In Artificial Life IV. Cambridge, Massachusetts.

[Siskind] Siskind, J.M. 1992. Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

[Stein 94] Stein, L.A. 1994. Imagination as Situated Cognition. Journal of Experimental and Theoretical Intelligence 6:393–407.

[Stein 96] Stein, L.A. 1996. Neo-Modular Systems: Architectural Principles for Cognitive Robotics. AAAI Fall Symposium on Embodied Cognition and Action. Cambridge, Massachusetts.

[Sutton] Sutton, R.S. 1990. Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming. In Proceedings of the Seventh International Conference on Machine Learning.

[Tesauro] Tesauro, G. 1995. Temporal Difference Learning and TD-Gammon. Communications of the ACM 38(3).

[Torrance] Torrance, M.C. 1994. Natural Communication with Robots. Sc.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

[Wessler] Wessler, M. 1995. A Modular Visual Tracking System. Sc.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

[Williamson] Williamson, M.M. 1996. Postural Primitives: Intteractive Behavior for a Humanoid Robot Arm. In Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior. Cape Cod, Massachusetts.

[Yanco and Stein] Yanco, H.; and L. A. Stein. 1992. An Adaptive Communication Protocol for Cooperating Mobile Robots. In Proceedings of the Second International Conference on the Simulation of Adaptive Behavior. Honolulu, Hawaii.

[Yanco] Yanco, Holly A. 1994. Robot Communication: Issues and Implementation. Sc.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.