Title
The Engineering of Physical Grounding

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

Author
Brooks, Rodney A.

Publication Date
1993

Peer reviewed

The Engineering of Physical Grounding

Rodney A. Brooks
MIT Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139, USA
brooks@ai.mit.edu

Over the last eight years there has been a renewed interest in building experimental mobile robot systems that operate in unadorned, and unmodified natural and unstructured environments. The act of building and programming such robots has had a significant impact since it has changed our understanding of what is difficult and what is easy. We have realized that symbolic systems for reasoning and planning are not sufficient for building embodied mobile robots situated in dynamic worlds. The new work has fundamentally changed planning research to have to deal with "reactive planning", has changed the emphasis in computer vision from recovery from single images or canned sequences of images to active, or animate, vision where the observer is a participant in the world controlling the imaging process in order to simplify the processing requirements, and has shifted some emphasis from centralized to distributed systems.

There is a general acceptance of the importance of subsymbolic lower levels that provide a reactive component for such embodied robots. But classical cognitive science still feels rather secure—things have only changed at the periphery. A new round of physical experimentation has the potential to challenge this orthodoxy.

The idea is to leap from experimenting with mobile robot systems to an almost humanoid-like integrated head system with saccading foveated vision, facilities for sound processing and sound production, and a compliant, dextrous manipulator. Computation would be handled by a large number of processors dedicated to particular subfunctions, and interconnected by a fixed topology network.

It is practical to build such a physical system today, although it is arguable that it could not be done only two years ago.

The question of how it should be programmed remains. If we apply the same techniques as was used in programming mobile robots, we have at the very least the chance of making a hand-eye system that is significantly more robust in an unstructured environment than any system that has been demonstrated in a laboratory to date.

Eight years ago mobile robots moved at agonizingly slow speeds of 1 meter every fifteen minutes and really did not appear to an external observer to understand their physical surroundings in any deep way. Today, using less computation, mobile robots are able to navigate about using purely visual guidance at 1 meter per second, giving tours of the MIT AI Lab. Unfortunately, hand-eye systems are still in the same dark ages inhabited by mobile robots eight years ago. Only in carefully controlled industrial settings do robots arms move with any speed grasping and manipulating objects. As soon as there is a small deviation from the expected state of the world however, such industrial robots fail completely. Laboratory systems working in an outwardly appearing less structured world painstakingly plan in a supposedly complete world model built from sensory input and a tiny library of possible objects. They take minutes to sense the world, plan, and grasp in a way which appears spastic at best. By applying the same ideas of layering competences each physically grounded there is hope that we could build a much more dextrous system able to operate on timescales commensurate with those of a person engaged in hand-eye coordination.

But there is a chance of a much richer set of outcomes. Such a physical system will let us test some ideas in physical reality that have until this point been at best philosophical musings.

The idea would be to take seriously the idea of the body in the mind, where everything is grounded in primitive sensor-motor patterns of activation. It would be a model above the neural level, in the sense that there would be no attempt to claim that the networks built had any direct biological validity, but below the usual cognitive level where manipulable symbols are assumed. The networks then would in principle share many of the properties of wet neural systems; completely distributed, highly interconnected, non-symbolic content to their messages, etc. Given the physical hardware a large scale set of programs would be built to allow the system to interact with people, as it watches them come and go from its environment, and manipulate objects within its reach.

Here are some of the challenges of such an approach:

• Can we have the system build "schemas" that are physically grounded in sensor-motor activation patterns while maintaining neuroscientific validity?

• Can such schemas really be used to construct higher levelabstractions? That are useful for visual recognition?

• Can such higher level abstractions be built at all in terms of lower level activation patterns without resorting to the meta-objectification which would seem to be the hallmark of mentalese or, more benignly, language.

• Can the spatial reasoning necessary to manipulate objects in the world be constructed out of re-used topographic image maps?

• Can such spatial reasoning systems be re-used to implement meta-objectification (the act of making a symbol), and as such what sources of constraint does that imply for language?

None of these questions can be tackled trivially. Each will take major implementation efforts. But the point is that the technology now exists to do such experiments. All that is needed is the intellectual energy, patience, and dedication, to carry them through.

The act of building and programming such a system would force us to face the issues of not only the interfaces between traditionally assumed modularities, but even the modularity itself. The issues that would have to be faced would strike at the very core of traditional Artificial Intelligence, cognitive science and linguistics, stressing theories there in ways that they are not normally stressed by standard operation within those fields. By reaching across standard boundaries and seriously attempting to get many sensing and acting modalities working together we may well quickly illuminate shortcomings in the standard models, shedding light on formerly unrealized sociologically shared, but wrong, assumptions. This is exactly the experience gained from work with mobile robots—earlier work simply was not asking the right questions, but assumed the standard intellectual decompositions could be put back together into working complete systems.

This approach to the grounding problem is an engineering approach. Build a system based on some hypotheses of physical grounding, and see how much further it can be taken than existing systems based on the physical symbol system hypothesis.