The relationship between **matter and life**

Rodney Brooks

The disciplines of artificial intelligence and artificial life build computational systems inspired by various aspects of life. Despite the fact that living systems are composed only of non-living atoms there seems to be limits in the current levels of understanding within these disciplines in what is necessary to bridge the gap between non-living and living matter.

Researchers in artificial intelligence (AI) and artificial life (Alife) are interested in understanding the properties of living organisms so that they can build artificial systems that exhibit these properties for useful purposes. AI researchers are interested mostly in perception, cognition and generation of action (Box 1), whereas Alife focuses on evolution, reproduction, morphogenesis and metabolism (Box 2). Neither of these disciplines is a conventional science; rather, they are a mixture of science and engineering. Despite, or perhaps because of, this hybrid structure, both disciplines have been very successful and our world is full of their products.

Every time we use a computer we use algorithms and techniques developed by AI researchers. These range from the natural language processing and indexing techniques in web search engines to the bayesian matching techniques used in help and document autoformatting systems in our word processors. When we play a video game our opponent is usually an AI system. At many airports, an AI program schedules our arrival gate, and when we apply for credit an AI neural network often vets our application. When we watch a film with digitally generated crowds, be they aliens or ants, we are watching groups of agents acting under Alife models of group behaviour. When we fly in the latest aeroplane, the design of the turbines may have been optimized by artificial evolution.

But despite all this, both fields have been labelled as failures for not having lived up to grandiose promises. At the heart of this disappointment lies the fact that neither AI nor Alife has produced artefacts that could be confused with a living organism for more than an instant. AI just does not seem as present or aware as even a simple animal and Alife cannot match the complexities of the simplest forms of life.

Moore power...

Part of the problem was the lack of computer Moore's law states that computational resources for a fixed price roughly double every 18 months. From about 1975 into the early 1990s all the gains of Moore's law went into the changeover from the centralized mainframe to the individual computer on your desk, accommodating a vastly increased number of users. The amount of computing power available to the individual scientist did not change that much, although the price came down by a factor of a thousand. But since the early 1990s, all of Moore's law has gone into increasing the performance of the workstation itself. And both AI and Alife have benefited from this shift.

Increased computer power has enabled search-based AI to push ahead with programs that achieve their ends through brute force — the Deep Blue program that beat the world chess champion is a good example. The essential ideas were in place in Greenblatt's 1965 program MacHack[1], but this could process only a few thousand possible chess moves per second. By 1997, when Deep Blue beat Kasparov, it was processing 200 million moves per second. More power has also enabled implementation of real-time perceptual systems, often based on neural models, that can simulate in serial computers the massive parallelism found in the brains of animals. Marr and Hildreth[2] required 10 minutes of computer time to find the edges in a single image in the late 1970s; we now have computer vision systems that track multiple moving objects in a scene at 30 frames per second. Others can visually track the boundaries of roads and cars a few times per second. Using such a system, the 'No hands across America'[3] project, at Carnegie Mellon University, made an automated truck drive from the east to the west coast of the United States with no human control for 98% of the journey.

Much more complex systems[4] can now be modelled as Alife, down to the level of molecules and enzyme-like components interact- complexity of the models is still far below that of any living system. New experiments in evolution simulate spatially isolated populations to investigate speciation. Over the past few years, new directions have emerged in AI[5], in attempts to implement artificial creatures in simulated or physical environments.

Often called the behaviour-based approach, this new mode of thought involves the connection of perception to action with little in the way of intervening representational systems. Rather than relying on search, this approach relies on the correct short, fast connections being present between sensory and motor modules. Behaviour-based approaches began with insect models, but more recently they have been extended to humanoid robots[6] — robots with human form that can interact with people in a social manner, eliciting natural and involuntary social responses from naive subjects.

Behaviour-based systems are now making their mark in consumer products such as the new generation of intelligent dolls, but perhaps their greatest success was the Sojourner Mars rover[7]. For the final phase of its mission in 1997 this behaviour-based robot was allowed to operate autonomously, and successfully navigated the surface of Mars.

The behaviour-based approach to AI has merged somewhat with the Alife endeavour, and a community of researchers has formed that is separate from the traditional AI community. The former is interested in understanding how living systems work and in building computational and physical models of them. The latter is interested in building systems with maximal performance, and is usually wary of biological inspiration as taking away from mathematically optimized engineering solutions.

Problems, problems, problems

Although they are much more lifelike than the pure engineering artefacts of traditional AI, in some sense the systems built under the

not seem as alive as we might hope. We build models to understand the biological systems better, but the models never work as well as biology. We have become very good at modelling fluids, materials, planetary dynamics, nuclear explosions and all manner of physical systems. Put some parameters into a program, let it crank, and out come accurate predictions of the physical character of the modelled system. But we are not good at modelling living systems, at small or large scales. Something is wrong.

Solutions and new developments?

What is going wrong? There are a number of possibilities: (1) we might just be getting a few parameters wrong; (2) we might be building models that are below some complexity threshold; (3) perhaps it is still a lack of computing power; and (4) we might be missing something fundamental and currently unimagined in our models of biology.

Incorrect parameters

Getting just a few parameters wrong would mean that we have essentially modelled everything correctly, but are just unlucky or ignorant in some minor way. With a bit more work on our part, things will start working better. It could be that our current neural-network models will work quantitatively better if we have five layers of artificial neurons, rather than today's standard of three. Or that artificial evolution works much better with populations of 100,000 or more, rather than the typical thousand or less. But this seems unlikely. One would expect that someone would have stumbled by now across a combination of parameters that worked qualitatively better than anything else around. That success would have led to theoretical analysis and we would have already seen rapid progress.

Models lack complexity

Building models that are below some complexity threshold also would mean that there is nothing in principle that we do not understand about intelligent or living systems. We have all the ideas and components lying around, we just have not yet put enough of them together in one place, or one model. When, and if, we do, then everything will start working a lot better. As for the first possibility, while this may be true, it does seem unlikely that is true across so many different aspects of biology.

A lack of computing power

We have recently seen an example of this. After being defeated by Deep Blue, Garry Kasparov said that he was surprised by its "ability to play as though it had a plan and how it understood the essence of the position". From the earlier versions he had been playing in the late 1980s. Deep Blue still had no strategic planning phase, as other chess programs designed to model human playing had. It still had only a tactical search, albeit a very deep, fast tactical search. This appeared to Kasparov to be about game plans, not because there was anything new, but because

Box 1
Artificial intelligence

Artificial intelligence (AI) received its name at a workshop[10] held by John McCarthy at Dartmouth College in New Hampshire in 1956. Marvin Minsky, Allen Newell and Herb Simon, together with John McCarthy, set the research agenda for machine intelligence for the next 30 years. All were inspired by earlier work by Alan Turing, Claude Shannon and Norbert Weiner on tree search for playing chess. From this workshop, tree search — for game playing, for proving theorems, for reasoning, for perceptual processes such as vision and speech and for learning — became the dominant mode of thought, and continues to be so today for large parts of the academic enterprise called AI. Search algorithms are mostly intrinsically serial, which contrasts with the ways in which large parts of animal nervous systems work, where there are both feedforward and feedback signals, but over very low-diameter networks. The speed of the individual neurons is so slow compared with the speed of overall action of the creature that very different intrinsic computations must be at play. There have been numerous flirtations with neurally inspired networks, but most of these are used as component-learning black boxes within a more traditional framework of serial computation.

Game playing is perhaps the most visible success associated with search. Games from tic-tac-toe through to chess have been conquered by brute-force search. More elaborate models of how humans play games have been tried, but in each case performance has soon been overtaken by the unstoppable march of Moore's law. The game of Go, however, has so many possible moves that it has remained impervious to brute-force search. Whereas it seems unlikely that humans play chess by brute-force search, and investigation of methods of play other than search are attractive, Go enforces these investigations with a vengeance. Even when we have computers with the same level of processing power as the human brain, they will not be able to play a good game of Go using brute-force search alone.

Search, however, has also been successful in other areas. It is used in proving theorems, in mathematical manipulation systems such as Mathematica and Matlab, and in systems of speech understanding and natural language understanding.

feel qualitatively different. The same might happen to our models of intelligence and life, if we could only get enough computer power.

If any of the above is the case then we should expect great progress in AI and Alife as soon as someone stumbles across the things that need to be fixed. The details will not particularly surprise anyone, although the new developments will have great practical impact. They will lead to new insights in all the sciences that study living organisms, as they will give us new sorts of computer models with which we can test rafts of new hypotheses about how living systems operate.

Models lack unimagined features

But what if we are missing something fundamental and currently unimagined in our models? We would then need to find new ways of thinking about living systems to make any progress, and this will be much more disruptive to all biology. As an analogy, suppose we were building physical simulations of elastic objects falling and colliding. If we did not quite understand physics, we might leave out mass as a specifiable attribute of the objects. Their falling behaviour would at first seem correct, but as soon as we started to look at collisions we would notice that the physical world was not being modelled correctly.

So what might be the nature of this unimagined feature of life? One possibility is that some aspect of living systems is invisible to us right now. The current scientific view of living things is that they are machines whose components are biomolecules. It is not completely impossible that we might discover some new properties of biomolecules or some new ingredient. One might imagine something on a par with the discovery of X-rays a century ago, which eventually led to our still-evolving understanding of quantum mechanics. Relativity was the other such discovery of the twentieth century, and had a similarly disruptive impact on the basic understanding of physics. Some similar discovery might rock our understanding of the basis of living systems.

New stuff

Let us call this the 'new stuff' hypothesis — the hypothesis that there may be some extra sort of 'stuff' in living systems outside our current scientific understanding. Roger Penrose[8], for one, has already hypothesized a weak form of 'new stuff' as an explanation for consciousness. He suggests that quantum effects in the microtubules of nerve cells might be the locus of consciousness at the level of the individual cell, which combines in bigger wave functions at the organism level. Penrose has not worked out a real theory of how this might work. Rather, he has suggested that this may be a critical element

understanding. This is a weak form of new stuff because it does not rely on anything outside the realm of current physics. For some it may have a certain appeal in that it unifies a great discovery in physics with a great question in biology — the nature of consciousness. David Chalmers[9] has hypothesized a stronger form of new stuff as an alternative explanation for consciousness. He suggests that a fundamentally new type, of the order of importance of spin or charm in particle physics, say, may be necessary to explain consciousness. It would be a new sort of physical property of things in the Universe, subject to physical laws that we just do not yet understand. Other philosophers, both natural and religious, might hypothesize some more ineffable entity such as a soul or *elan vital* — the 'vital force'.

Another way that the unimaginable discovery might come about is through 'new mathematics'. This would not require any new physics to be present in living systems. We may simply not be seeing some fundamental mathematical description of what is going on in living systems and so be leaving it out of our AI and Alife models. What might this 'new mathematics' be? Candidates have included catastrophe theory, chaos theory, dynamical systems and wavelets. When each of these new mathematical techniques hit the market, researchers noticed ways in which they could be used to describe what is going on in living systems, and then tried to incorporate the same thing into their computational models. It is not clear whether the mathematical techniques in question are best used as descriptive tools or as generative components within the computational models. The latter approach seems at times misguided. However, none of these wonder techniques has really made the hoped-for improvements in our models. Looking at the physical nature of living systems, there seem to be certain mathematical properties that are not handled at all by any of these new techniques, or by any current model. One property is that the matter that makes up living systems obeys the laws of physics in ways that are expensive to simulate computationally. For instance, the membranes of cells have a shape determined by the continuous minimization of forces between molecules within the membrane and on either side of it. Another property is that matter does not simply appear and disappear in the physical world, but great care must be taken in a computational simulation to enforce this.

An analogy to the sort of thing that might be missing is computation — not as the undiscovered feature itself but as an analogy for the type of thing we might be looking for. For most of the twentieth century we have poked electrodes into living nervous systems signals measured and events that occur elsewhere in the creature. These data are used to test hypotheses about how the living system 'computes' in the broadest sense of the word. Imagine a society isolated for the past hundred years and in which computers have not been invented. If the scientists in this society came across a working computer, doing if they had no notion of computation? Would it make any sense without the notion of Turing computability, or an understanding of a Von Neumann architecture? Or would our isolated scientists need to reinvent the notion of computation before they could explain what the machine was doing? I strongly suspect that they would. Nothing that Turing or Von Neumann did in their mathematics at this level was particularly disruptive. A good late-nineteenth-century mathematician could understand it all with a few days instruction — there would be no surprises for them in the way that quantum mechanics and relativity would surprise a physicist from the same era.

So now we return to the unimaginable. For perceptual systems, say, there might be some organizing principle, some mathematical notion that we need in order to understand how they really work. If so, discovering this principle will enable us to build computer-vision systems that are good at separating objects from the background, understanding facial expression, discriminating the living from the non-living and general object recognition. None of our current vision systems can do much at all in any of these areas. What form might this mathematical notion take? It need not be disruptive of our current view of living things, but could be as non-threatening as the notion of computation, just different to anything anyone has currently thought of. Perhaps other mathematical principles or notions, necessary to build good explanations of the details of evolution, cognition, consciousness or learning, will be discovered or invented and let those subfields of AI and Alife flower. Or perhaps there will be just one mathematical notion, one 'new mathematics' idea, that will unify all these fields, revolutionize many aspects of research involving living systems, and enable rapid progress in AI and Alife. That would be surprising, delightful and exciting. And of course whether or not this will happen is totally unforeseeable.

Box 2 Artificial life

Chris Langton instigated the notion of artificial life (Alife) at a workshop[11] in Los Alamos, New Mexico, in 1987. The enterprise was to make living systems without the direct aid of biological structures. The work was inspired largely by John Von Neumann, and his early work on self-reproducing machines in cellular automata. Von Neumann's model of computation was a Turing machine, the same intrinsically serial mechanism adopted by AI. Researchers in Alife were much inspired by biological systems and quickly produced simulations and analytical models of aspects of biological reproduction and group behaviours of organisms.

Tom Ray[12] developed a system called Tierra, where multiple computer programs competed for the resource of the processing unit in a simulated computer. The seed program could reproduce itself, and took 80 nibbles (half-bytes) of code space. It followed the classical Von Neumann model, and had the central feature of molecular biology, in that the code needed to be both transcribed, or interpreted, and copied into the progeny so that it too had its own 'DNA'. The simulated computer, however, was subject to copying mistakes, and to 'cosmic rays' which randomly flipped bits in the memory space. Its memory filled up with copies of the original seed program, although some of them had errors and were removed. But others started to optimize and get smaller — there was environmental pressure selecting for small programs — and parasites less than half the size of the original programs soon evolved. Although not able to copy themselves, they could trick a larger program into copying them rather than itself. Diverse populations, such as one sees in nature, soon arose in the system.

Most recently, Lipson and Pollack[13] have evolved 'creatures' that can move. Their morphology is restricted to fixed-length bars with ball-and-socket joints. The designs evolved in a physical simulation program, and the fitness of individuals was determined by measuring their ability to move on a surface. Mutating the most successful members of the previous generation produced each new generation. The *tour de force* was to have the creatures automatically produced directly from the evolved genome by a computerized rapid prototyping system. Motors were snapped on by hand and the creatures, connected to a computer running the same neural network used in the simulations, then started crawling across physical surfaces.

Rodney Brooks is at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Massachusetts 02139, USA.

1. Greenblatt, R. D., Eastlake, D. E. & Crocker, S. D. in *Proc. Fall Joint Comput. Conf.* 801–810 (AFIPS Press, Montvale, NJ, 1967).
2. Marr, D. C. & Hildreth, E. *Proc. R. Soc. Lond. B* **207,** 187–217 (1980).
3. Pomerleau, D. A. & Jochem, T. M. *Photon. Spectra* 80–85 (April 1996).
4. Kitano, H. *New Generation Comput.* **18,** 199–216 (2000).
5. Brooks, R. A. *Science* **253,** 1227–1232 (1991).
6. Adams, B., Breazeal, C., Brooks, R. A. & Scassellati, B. *IEEE Intell. Syst. Applic.* **15,** 25–31 (2000).
7. Matijevic, J. *Science* **280,** 454–455 (1998).
8. Penrose, R. *Shadows of the Mind* (Oxford Univ. Press, New York, 1994).
9. Chalmers, D. *J. Conscious. Stud.* **2,** 200–219 (1995).
10. McCorduck, P. *Machines Who Think* (Freeman, New York, 1979).
11. Langton, C. G. (ed.) *Artificial Life* (Addison-Wesley, CA, 1989).
12. Ray, T. S. in *Artificial Life* Vol. II (ed. Langton, C. G.) 371–408 (Addison-Wesley, CA, 1991).