AI is possible .. but AI won't happen: The future of Artificial Intelligence

Dr. Mark Humphrys
formerly University of Edinburgh
now Dublin City University

A talk given to the Next Generation symposium,
Jesus College, Cambridge, Aug 1997.

Artificial Intelligence (AI) is a perfect example of how sometimes science moves more *slowly* than we would have predicted. In the first flush of enthusiasm at the invention of computers it was believed that we now finally had the tools with which to crack the problem of the mind, and within years we would see a new race of intelligent machines. We are older and wiser now. The first rush of enthusiasm is gone, the computers that impressed us so much back then do not impress us now, and we are soberly settling down to understand how hard the problems of AI really are.

What is AI? In some sense it is engineering inspired by biology. We look at animals, we look at humans and we want to be able to build machines that do what they do. We want machines to be able to learn in the way that they learn, to speak, to reason and eventually to have consciousness. AI is engineering but, at this stage, is it also science? Is it, for example, modelling in cognitive science? We would like to think that is both engineering and science but the contributions that is has made to cognitive science so far are perhaps weaker than the contributions that biology has given to the engineering.

The confused history of AI

Looking back at the history of AI, we can see that perhaps it began at the wrong end of the spectrum. If AI had been tackled logically, it would perhaps have begun as an *artificial biology*, looking at living things and saying "Can we model these with machines?". The working hypothesis would have been that living things are physical systems so let's try and see where the modelling takes us and where it breaks down. Artificial biology would look at the evolution of physical systems in general, development from infant to adult, self-organisation, complexity and so on. Then, as a subfield of that, a sort of *artificial zoology* that looks at sensorimotor behaviour, vision and navigation, recognising, avoiding and manipulating objects, basic, pre-linguistic learning and planning, and the simplest forms of internal representations of external objects. And finally, as a further subfield of this, an *artificial psychology* that looks at human behaviour where we deal with abstract reasoning, language, speech and social culture, and all those philosophical conundrums like consciousness, free will and so forth.

That would have been a logical progression and is what should have happened. But what did happen was that what people thought of as intelligence was the stuff that *impresses* us. Our peers are impressed by things like doing complex mathematics and playing a good chess game. The ability to walk, in contrast, doesn't impress anyone. You can't say to your friends, "Look, I can walk", because your friends can walk too.

So all those problems that toddlers grapple with every day were seen as unglamorous, boring, and probably pretty easy anyway. The really *hard* problems, clearly, were things demanding abstract thought, like chess and mathematical theorem proving. Everyone ignored the animal and went straight to the human, and the adult human too, not even the child human. And this is what `AI' has come to mean - artificial adult human intelligence. But what has happened over the last 40-50 years - to the disappointment of all those who made breathless predictions about where AI would go - is that things such as playing chess have turned out to be incredibly easy for computers, whereas learning to walk and learning to get around in the world without falling over has proved to be unbelievably difficult.

And it is not as if we can ignore the latter skills and just carry on with human-level AI. It has proved very difficult to endow machines with `common sense', emotions and those other intangibles which seem to drive much intelligent human behaviour, and it does seem that these may come more from our long history of interactions with the world and other humans than from any abstract reasoning and logical deduction. That is, the animal and child levels may be the *key* to making really convincing, well-rounded forms of intelligence, rather than the intelligence of chess-playing machines like Deep Blue, which are too easy to dismiss as `mindless'.

In retrospect, the new view makes sense. It took 3 billion years of evolution to produce apes, and then only another 2 million years or so for languages and all the things that we are impressed by to appear. That's perhaps an indication that once you've got the mobile, tactile monkey, once you've got the *Homo erectus*, those human skills can evolve fairly quickly. It may be a fairly trivial matter for language and reasoning to evolve in a creature which can already find its way around the world.

The new AI, and the new optimism

That's certainly what the history of AI has served to bear out. As a result, there has been a revolution in the field which goes by names such as Artificial Life (AL) and Adaptive Behavior, trying to re-situate AI within the context of an artificial biology and zoology (respectively). The basic philosophy is that we need much more understanding of the animal substrates of human behaviour before we can fulfil the dreams of AI in replicating convincing well-rounded intelligence.

(Incidentally, the reader should note that the terminology is in chaos, as fields re-group and re-define themselves. For example, I work on artificial zoology but describe myself casually as doing AI. This chaos can, however, be seen as a healthy sign of a field

which has not yet stabilised. Any young scientist with imagination should realise that these are the kind of fields to get into. Who wants to be in a field where everything was solved long ago?)

So AI is not dead, but re-grouping, and is still being driven, as always, by testable scientific models. Discussions on philosophical questions, such as `What is life?' or `What is intelligence?', change little over the years. There have been numerous attempts, from Roger Penrose to Gerald Edelman, to disprove AI (show that it is impossible) but none of these attempted revolutions has yet gathered much momentum. This is not just because of lack of agreement with their philosophical analysis (although there is plenty of that), but also perhaps because they fail to provide an alternative paradigm in which we can do science. Progress, as is normal in science, comes from building things and running experiments, and the flow of new and strange machines from AI laboratories is not remotely exhausted. On the contrary, it has been recently invigorated by the new biological approach.

In fact, the old optimism has even been resurrected. Professor Kevin Warwick of the University of Reading has recently predicted that the new approach will lead to human-level AI in our lifetimes. But I think we have learned our lesson on that one. I, and many like me in new AI, imagine that this is still Physics before Newton, that the field might have a good one or two hundred years left to run. The reason is that there is no obvious way of getting from here to there - to human-level intelligence from the rather useless robots and brittle software programs that we have nowadays. A long series of conceptual breakthroughs are needed, and this kind of thinking is very difficult to timetable. What we are trying to do in the next generation is essentially to find out what are the right questions to ask.

It may never happen (but not for the reasons you think)

I think that people who are worried about robots taking over the world should go to a robotics conference and watch these things try to walk. They fall over, bump into walls and end up with their legs thrashing or wheels spinning in the air. I'm told that in this summer's Robotic Football competition, the *losing* player scored all five goals - 2 against the opposing robot, and 3 against himself. The winner presumably just fell over.

Robots are more helpless than threatening. They are really quite sweet. I was in the MIT robotics laboratory once looking at Cog, Rodney Brooks' latest robot. Poor Cog has no legs. He is a sort of humanoid, a torso stuck on a stand with arms, grippers, binocular vision and so on. I saw Cog on a Sunday afternoon in a darkened laboratory when everyone had gone home and I felt sorry for him which I know is mad. But it was Sunday afternoon and no one was going to come and play with him. If you consider the gulf between that and what most animals experience in their lives, surrounded by a tribe of fellow infants and adults, growing up with parents who are constantly with them and constantly stimulating them, then you understand the incredibly limited kind of life that artificial systems have.

The argument I am developing is that there may be limits to AI, not because the hypothesis of `strong AI' is false, but for more mundane reasons. The argument, which I develop further on my website, is that you can't expect to build single isolated AI's, alone in laboratories, and get anywhere. Unless the creatures can have the space in which to evolve a rich culture, with repeated social interaction with things that are like them, you can't really expect to get beyond a certain stage. If we work up from insects to dogs to *Homo erectus* to humans, the AI project will I claim fall apart somewhere around the *Homo erectus* stage because of our inability to provide them with a real cultural environment. We cannot make millions of these things and give them the living space in which to develop their own primitive societies, language and cultures. We can't because the planet is already full. That's the main argument, and the reason for the title of this talk.

So what will happen?

So what *will* happen? What will happen over the next thirty years is that will see new types of animal-inspired machines that are more `messy' and unpredictable than any we have seen before. These machines will change over time as a result of their interactions with us and with the world. These silent, pre-linguistic, animal-like machines will be nothing like humans but they will gradually come to seem like a strange sort of animal. Machines that learn, familiar to researchers in labs for many years, will finally become mainstream and enter the public consciousness.

What category of problems could animal-like machines address? The kind of problems we are going to see this approach tackle will be problems that are somewhat noise and error resistant and that do not demand abstract reasoning. A special focus will be behaviour that is easier to *learn* than to articulate - most of us know how to walk but we couldn't possibly tell anyone how we do it. Similarly with grasping objects and other such skills. These things involve building neural networks, filling in state-spaces and so on, and cannot be captured as a set of rules that we speak in language. You must experience the dynamics of your own body in infancy and thrash about until the changing internal numbers and weights start to converge on the correct behaviour. Different bodies mean different dynamics. And robots that can learn to walk can learn other sensorimotor skills that we can neither articulate *nor* perform ourselves.

What are examples of these type of problems? Well, for example, there are already autonomous lawnmowers that will wander around gardens all afternoon. The next step might be autonomous vacuum cleaners inside the house (though clutter and stairs present immediate problems for wheeled robots). These are all sorts of other uses for artificial animals in areas where people find jobs dangerous or tedious - land-mine clearance, toxic waste clearance, farming, mining, demolition, finding objects and robotic exploration, for example. Any jobs done currently or traditionally by animals would be a focus. We are familiar already from the Mars Pathfinder and other examples that we can send autonomous robots not only to inhospitable places, but also send them

there on cheap one-way `suicide' missions. (Of course, no machine ever `dies', since we can restore its mind in a new body on earth after the mission.)

Whether these type of machines may have a future in the home is an interesting question. If it ever happens, I think it will be because the robot is treated as a kind of pet, so that a machine roaming the house is regarded as cute rather than creepy. Machines that learn tend to develop an individual, unrepeatable character which humans can find quite attractive. There are already a few games in software - such as the Windows-based game Creatures, and the little Tamagotchi toys - whose personalities people can get very attached to. A major part of the appeal is the unique, fragile and unrepeatable nature of the software beings you interact with. If your Creature dies, you may never be able to raise another one like it again. Machines in the future will be similar, and the family robot will after a few years be, like a pet, literally irreplaceable.

What will hold things up? There are many things that could hold up progress but hardware is the one that is staring us in the face at the moment. Nobody is going to buy a robotic vacuum cleaner that costs £5000 no matter how many big cute eyes are painted on it or even if it has a voice that says, "I love you". Many *conceptual* breakthroughs will be needed to create artificial animals. The major theoretical issue to be solved is probably representation: what is language and how do we classify the world. We say `That's a table' and so on for different objects, but what does an insect do, what is going on in an insect's head when it distinguishes objects in the world, what information is being passed around inside, what kind of data structures are they using. Each robot will have to learn an internal language customised for its sensorimotor system and the particular environmental niche in which it finds itself. It will have to learn this internal language on its own, since any representations we attempt to impose on it, coming from a different sensorimotor world, will probably not work.

Predictions

Finally, what will be the impact on society of animal-like machines? Let's make a few predictions that I will later look back and laugh at.

First, family robots may be permanently connected to wireless family intranets, sharing information with those who you want to know where you are. You may never need to worry if your loved ones are alright when they are late or far away, because you will be permanently connected to them. Crime may get difficult if all family homes are full of half-aware, loyal family machines. In the future, we may never be entirely alone, and if the controls are in the hands of our loved ones rather than the state, that may not be such a bad thing.

Slightly further ahead, if some of the intelligence of the horse can be put back into the automobile, thousands of lives could be saved, as cars become nervous of their drunk owners, and refuse to get into positions where they would crash at high speed. We may

look back in amazement at the carnage tolerated in this age, when every western country had road deaths equivalent to a long, slow-burning war. In the future, drunks will be able to use cars, which will take them home like loyal horses. And not just drunks, but children, the old and infirm, the blind, all will be empowered.

Eventually, if cars were all (wireless) networked, and humans stopped driving altogether, we might scrap the vast amount of clutter all over our road system - signposts, markings, traffic lights, roundabouts, central reservations - and return our roads to a soft, sparse, eighteenth-century look. All the information - negotiation with other cars, traffic and route updates - would come over the network invisibly. And our towns and countryside would look so much sparser and more peaceful.

Conclusion

I've been trying to give an idea of how artificial animals could be useful, but the reason that I'm interested in them is the hope that artificial animals will provide the route to artificial humans. But the latter is not going to happen in our lifetimes (and indeed may never happen, at least not in any straightforward way).

In the coming decades, we shouldn't expect that the human race will become extinct and be replaced by robots. We can expect that classical AI will go on producing more and more sophisticated applications in restricted domains - expert systems, chess programs, Internet agents - but any time we expect common sense we will continue to be disappointed as we have been in the past. At vulnerable points these will continue to be exposed as `blind automata'. Whereas animal-based AI or AL will go on producing stranger and stranger machines, less rationally intelligent but more rounded and whole, in which we will start to feel that there is somebody at home, in a strange animal kind of way. In conclusion, we won't see full AI in our lives, but we should live to get a good feel for whether or not it is possible, and how it could be achieved by our descendants.

Further Reading

- *Darwin's Dangerous Idea*, Daniel C. Dennett (philosophical background). Dennett shows how Strong AI is simply the consequence of ordinary scientific materialism, and any alternative better fit into evolutionary materialism as well as AI does.
- *Out of Control: The New Biology of Machines*, Kevin Kelly. A wonderfully written survey of current work.
- *Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 years*, Jared Diamond (evolutionary history). Diamond demonstrates vividly how easily cultures fail, and how hard our human success was. AIs will be even more vulnerable to cultural failure.

Return to my home page.

humphrys@compapp.dcu.ie