

The Inevitable Magic of Artificial Life

Hod Lipson, Cornell University

From: Pfeifer R., (Ed.) *The Rediscovery of Intelligence*, 2007, pp. 114-115

Nothing is more exciting to a passionate Robotics or Artificial Life researcher than the thought that their creation will, eventually, be capable of sustaining itself to a degree that qualifies as a real form of life. Where does this passion come from? Are we all the modern incarnation of those ancient wizards who tried to breathe life into inanimate matter? Today, with perhaps a few more tools in our pocket, possibly slightly better understanding, and sometimes a little more funding, we are still trying.

We are, of course, well aware of the potential dangers. From ancient myths such as the Jewish Golem legend through children's tales as *The Sorcerer's Apprentice*, warnings abound about the consequences of taking god's power into one's own hands. Their modern incarnation in Hollywood movies tell similar stories with an ever increasing level of details. All tales end in some sort of a disaster, from machines killing their creators, if not taking over the world. Human enslavement is inevitable, as the cold machines take control. But despite virtually all these romantic depictions ending in tragedy, we persist. Even as the robots take over, we will rejoice in our own success, and in our last living moments we will crawl over to the computer to hit the enter key for one last time, to send our manuscript off for publication.

Or will we?

Indeed the power of creation, from art to technology, is seductive, and creating life would no doubt rank as an ultimate triumph. Not only would it be a triumph of creation (engineering), it would be a triumph of understanding (science) as well, because in order to have achieved creation, the underlying fundamental processes must have been teased out completely.

Unless, of course, we didn't actually create it ourselves, but instead we have evoked something else to create it for us – another machine, perhaps?

I often find myself contemplating this point with colleagues who take the view that the humans reserve a unique place in nature that fundamentally cannot be emulated by a machine. We usually agree, fairly quickly, that it is not a question of technology: At the current rate of progress, it is only a matter of time until the ability to *emulate* a human (or a pet) will be within reach. In 20 years? In 100 years? Surely in a thousand years. Eventually – we'll get there, technologically. But would humans *buy into it*? Will it really be considered *authentic* if we know it's a machine? This is not the deeper question of *is it alive* or *what is human*, but a more pragmatic question of acceptance.

– “Would you *really* fall in love with a machine that was *made* by a human?” is a frequently evoked rhetoric. “Would you *really* feel any passion towards a machine that was *programmed* to say *I love you*?”

That argument makes sense: Our rational, reductionist, mechanical understanding of anything immediately demystifies it, removing any romantic mystique and leaving it naked and transparent.

The evidence, however, is sometimes surprising. People empathically anthropomorphize objects and devices: From their Tamaguchi keychains to their Aibo dog. People insist that when objects they love are sent for repair, they get back the exact one they sent, not any refurbished, “identical” replacement or even upgraded version. But I think that our ability to attach emotionally to these objects is not incidental, it is because they are things whose workings *we don't fully understand*.

And there lies the rub: The more we understand its inner workings, the less we are likely to emotionally appreciate it. People who become attached to an Aibo are typically not computer scientists or engineers.

So will we fall in love with a machine that has been programmed to say *I love you*? No, if indeed it was *programmed*. From the short history of AI, however, it is increasingly apparent that the ultimate mechanical life forms will not be designed *directly* by any engineer. They will not be programmed *directly* by any computer scientist. They will not be manufactured *directly* by a machinist. Instead, they will learn. They will evolve. They will grow. They will not be made, programmed, or designed. At all. Consequently, I believe, they *will* be lovable.

The future may thus be quite disappointing for us wannabe creators. If we haven't created it, can we justly claim the ultimate triumph of creation? In the early days of AI, it was somewhat exciting to see a new fact derived automatically without being directly programmed in. It was exciting to see a robot learn how to walk without directly programming it. But as modern AI progresses, we resort more and more to indirect methods – shifting gradually from direct and symbolic learning methods to more abstract, black-box machine learning processes where the resulting model is not explicitly understood. More recently, we use stochastic learning processes such as evolutionary search, where the adaptation process itself is not well understood.

And so, creating the ultimate robotic lifeform will likely be no more revealing, but no less awe inspiring, than planting a seed and watching it grow into a tree. Like a child, we may be able only to indirectly influence the development and growth of a robot by controlling its environment and experiences, but we will not be directly involved in its design, manufacture or programming more than we are in any living creature. Robots will be complicated and difficult to understand and predict, their inner workings concealed by a barrier of complexity. They will exploit a mix of connectionist, developmental and morphological computation processes that will require a new type of biology-like science dedicated to reverse-engineering them. That complexity barrier will retain a degree of enchantment that will permit emotional attachment, but at the same time, will prevent us from claiming the coveted feat of creation.

Will the enchantment of machines come at the expense of demystification of humans, causing us to see each other in new light? Perhaps, and even if we don't learn much about the robots in the end, we might just learn something new about ourselves.