Ethics and artificial life: From modeling to moral agents

John P. Sullins
Philosophy Department, Sonoma State University, 1801 East Cotati Avenue, Rohnert Park, CA 94928-3609, USA
E-mail: John.sullins@Sonoma.edu

Abstract.

Artificial Life (ALife) has two goals. One attempts to describe fundamental qualities of living systems through agent based computer models. And the second studies whether or not we can artificially create living things in computational mediums that can be realized either, virtually in software, or through biotechnology. The study of ALife has recently branched into two further subdivisions, one is ''dry'' ALife, which is the study of living systems ''in silico'' through the use of computer simulations, and the other is ''wet'' ALife that uses biological material to realize what has only been simulated on computers, effectively wet ALife uses biological material as a kind of computer. This is challenging to the field of computer ethics as it points towards a future in which computer and bioethics might have shared concerns. The emerging studies into wet ALife are likely to provide strong empirical evidence for ALife's most challenging hypothesis: that life is a certain set of computable functions that can be duplicated in any medium. I believe this will propel ALife into the midst of the mother of all cultural battles that has been gathering around the emergence of biotechnology. Philosophers need to pay close attention to this debate and can serve a vital role in clarifying and resolving the dispute. But even if ALife is merely a computer modeling technique that sheds light on living systems, it still has a number of significant ethical implications such as its use in the modeling of moral and ethical systems, as well as in the creation of artificial moral agents.

Key words: artificial life, ethical status of artificial agents, machine ethics, simulating evolutionary ethics

Introduction

Artificial Life (ALife) has two goals. One attempts to describe fundamental qualities of living systems through agent based computer models. And the second studies whether or not we can artificially create living things in computational mediums that can be realized either, virtually in software, or through bio-technology. Strangely, except for the initial burst of articles and books on the subject when it was first introduced in the early 90s, [1]the exciting nature of this research has gone largely unnoticed by the general public and media. Philosophers have also paid insufficient attention to this subject and in particular have not helped work through the various ethical issues raised by the notion of creating living, or at least life-like, software agents or living

[1] S. Levy.Artificial Life: The Quest for a New Creation.
Pantheon Books, New York, 1992; R. Lewontin. Complexity: Life at the Edge of Chaos. Macmillan Publishing Company, New York, 1992; C. Emmeche.The Garden in the Machine, the Emerging Science of Artificial Life. Steven Sampson, translator, Princeton University Press, Princeton, 1994; J. Horgan. From Complexity to Perplexity.Scientific American, June: 104, 1995; J. Horgan. The End of Science. Addison-Wesley, Reading, 1996.

bio-computational entities. Reasonable arguments can be raised that ALife is not going to succeed in its most ambitious goals, but even if ALife is merely a computer modeling technique that sheds light on living systems, it still has a number of significant ethical implications that need to be addressed. The study of ALife has recently branched into two further subdivisions, one is ''dry'' ALife, which is the study of living systems ''in silico'' through the use of computer simulations and the other is ''wet'' ALife that uses biological material to realize what has only been simulated on computers, effectively wet ALife uses biological material as a kind of computer. This is challenging to the field of computer ethics as it points towards a future in which computer and bioethics might have shared concerns. The emerging studies into wet ALife are likely to provide strong empirical evidence for ALife's most challenging hypothesis: that life is a certain set of computable functions that can be duplicated in any medium. I believe this will propel ALife into the midst of the mother of all cultural battles that has been gathering around the emergence of biotechnology. Philosophers need to pay close attention to this debate and can serve a vital role in clarifying and resolving the dispute.

The unwelcome truths of ALife

Some early researchers in ALife made the claim that ALife can, or had already, synthesized real artificial life inside their computers.[2]Since that time, however, the vast majority of the researchers in this field have not taken a hard stand on this issue. While individual researchers may believe in their heart of hearts that their software agents are technically alive, they do not often posit that belief in their research. Instead, most researchers tend to use ALife as a fruitful modeling method with applications in a great variety of topics leaving the messy business of defining ''life'' to the philosophers and theoretical biologists. Perhaps this retreat from Hard ALife in favor of the more cautious Soft ALife stance is responsible for the loss of the public interest in the subject. The claim that ALife can provide a credible method for studying evolution in a software context is not at all as sexy as saying, ''my computer is alive.'' Others, have had some nagging doubts about Hard ALife,[3] Soft ALife has not raised much objection

[2] **2** C.G. Langton. Artificial Life. In Chris Langton, editor, SFI Studies in the Sciences of Complexity, Proc. Vol. VI. Addison-Wesley, Redwood City, 1989; S. Rasmussen. Aspects of Information, Life, Reality, and Physics. In C. Langton, C. Taylor, J.D. Farmer, S. Rasmussen, editors. Artificial Life II: the Proceedings of the Workshop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10, pp. 767–774. Addison-Wesley, Redwood City, 1992; T. Ray. An Approach to the Synthesis of Life. In C. Langton, C. Taylor, J.D. Farmer, S. Rasmussen, editors. Artificial Life II: the Proceedings of the Workshop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10, pp. 767–774. Addison-Wesley, Redwood City, 1992.

[3] **3** M.A. Bedeau. Philosophical Aspects of Artificial Life.
In Towards a Practice of Autonomous Systems. Proceedings of the first European Conference on Artificial Life, pp. 494–503. MIT Press, Cambridge, 1992; Emmeche, 1994; Horgan, 1996; E.T. Olson. The Ontological Basis of Strong Artificial Life. Artificial Life, 3(1): 29–39, 1997; H.H. Pattee. Simulations, Realizations, and Theories of Life. In C. Langton, editor, Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living

from philosophers. However, I would like to argue here that both strong and weak ALife has the potential to be seen as a threat to some traditional thinkers because ALife, strong or weak, provides compelling evidence for the power of evolution and its role in all life, including that of humans, and this has been a hot topic to traditional thinkers, especially in America.

Let me explain why I feel this is an important topic for discussion in the field of computer ethics. Computing technology has consistently challenged many of our traditional ethical values. Navigating the rapidly shifting landscape of computing technology as well as analyzing the computer's role in our ethical and belief systems have long been the purview of the field of computer ethics. For example, the ease in which one can access and process data with a computer has led to new challenges in maintaining personal privacy that did not pose much of a threat to earlier generations. So the technology itself exacerbates existing ethical problems and in some situations creates entirely new ones. In these instances the computer is a new tool through which human agents impact one another and these altered relationships can be evaluated from an ethical standpoint. This general class of cases and conundrums makes up the core of computer ethics study but there are two additional ways in which the computer can effect ethical deliberations.

One is that the computer can serve as a metaphor or model that alters the way we look at a subject and in that way alter or challenge traditional values and mores. For instance, evolution is a difficult concept for the non-specialist to grasp, but with a computer and a few ALife models it can be quickly and effectively taught. For instance, Richard Dawkins in his book; The Blind Watchmaker,[4] uses a simple ALife model to prove his point that one can get complex phenotypic structures from the simple random mutation of a set of ''genes'' that are encoded in the software and subjected to the fitness function of pleasing the aesthetics of the user. This program results in interesting patterns that develop without any preset design for that structure in the mind of the user.[5] Dawkins'

Systems, held September 1987, in Los Alamos, New Mexico, Vol. 6. Addison-Wesley, Redwood City, CA, 1989; E. Sober. Learning From Functionalism – Prospects for Strong Artificial Life. In C. Langton, et al., editors, Artificial Life II: The Proceedings of the Workshop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10. Addison-Wesley, Redwood City, CA, 1992; J. Sullins. Knowing Life: Possible Solutions to the Practical Epistemological Limits in the Study of Artificial Life. The Journal of Experimental and Theoretical Artificial Intelligence, 13 (4), pp. 397–408, 2001.

[4] R. Dawkins.The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design. W.W. Norton and Company, Inc., 1996.

[5] Here is a similar java program that can be accessed from the web and experimented with by the reader: (http:// suhep.phy.syr.edu/courses/mirror/biomorph/). have on the study of ethics, human nature and morality. The compelling illustrations presented by these software models in favor of an evolutionary understanding of our world, and our own human nature, will necessarily place ALife in direct confrontation with those who would prefer to excise all discussion of evolution in the context of ethics and human nature.

program is relatively simple but other, more complex, programs exist and through these ALife programs one can learn the dynamics of the theory of evolution, including its beauty and its ugliness, and the simplicity of its components that nonetheless have the ability to produce complex organisms and behaviors. Later in this paper we will look at a few more examples of these kinds of models and discuss the impacts they might

The second alternative way in which computer technology can impact our ethics is when it plays the role of an independent moral agent. This is still just a theoretical possibility, but it is not inconceivable that computer software agents or advanced physical robots might someday become robust enough in their dealings with humans and each other that they will have to be considered as part of our moral deliberations.[6] If one believes in the hard ALife claim that some computer agents are, or could conceivable become, living agents with their own interests, environments and ecologies, then there is a real possibility that these theoretical entities might have to be considered from an ethical standpoint. The arguments in favor of this ethical stance could proceed in the same way that proponents of animal rights might extend moral concern to animals, or possibly one might see a significantly advanced ALife ecosystem as a kind of natural environment that deserves the kind of regard that environmental ethics argues we must pay towards natural biological ecosystems.

So there are two areas that need to be address in this paper; the ethics of ALife as models (Soft ALife), and the ethics of ALife as independent agents and ecosystems (Hard ALife). We will find that in both areas there are some unwelcome findings that challenge traditional ethical systems.

Ethics and soft ALife

It is easy to see that certain aspects of soft ALife would be of interest to traditional computer ethics concerns. For instance, soft ALife deals entirely with self-replicating programs, a trait that is shared by those who create computer viruses. So there are a number of issues that can be raised regarding the use and containment of legitimate ALife programs, which might become a real problem if they were used maliciously.

In this paper though, I am interested in a more subtle concern. The computer is a powerful medium for realizing complex thought experiments. For instance, imagining the process and hypothetical mechanisms of the theory of evolution is a difficult problem that is vastly simplified when mathematical and computational tools are used to help illustrate it. Darwin's theory is simple to state but vastly counterintuitive, so much so that it still does not find

---

[6] M. Anderson, S.L. Anderson, and C. Armen. Towards Machine Ethics.Proceedings of AAAI Workshop on Agent Organizations: Theory and Practice, San Jose, CA, July, 2004; M. Elton. Should Vegetarians Play Video Games? Philosophical Papers, 2000.

acceptance with traditional thinkers. As Daniel Dennett explains in his book; Darwin's Dangerous Idea:

Here then is Darwin's dangerous idea: the algorithmic levelis the level that best accounts for the speed of the antelope, the wing of an eagle, the shape of the orchid, the diversity of species, and all the other occasions for wonder in the world of nature. It is hard to believe that something as mindless and mechanical as an algorithm could produce such wonderful things.[7]

Understanding the power of algorithms is impossible without having some experience in manipulating them and watching them work. The algorithm of evolution is further obscured by the fact that it is a recursive algorithm that moves so slowly one cannot observe it easily in nature. ALife models remove both of these difficulties by allowing one to build simulations where the user can modify fitness constraints, mutation rates, etc. and then run through millions of generations of software entities in a few seconds, which makes soft ALife programs a compelling way to illustrate the veracity of Darwin's dangerous idea, and as such, these models could be seen to be dangerous as well.

One common argument used by those antagonistic to the theory of evolution is that the belief in evolution corrupts our faith in traditional ethics, since without an intelligent designer there is no grounding to our moral intuitions and some sort of Hobbesian state of nature will replace our well-ordered societies. I would like to take a brief look at a few ALife programs to counter that argument, by showing how societies and even altruism can be a product of evolutionary forces and since altruism is an other-directed behavior, it is conceivable that more complex other-directed behaviors such as ethics and morality might develop under evolutionary constraints as well.

Existing ALife models that shed light on the evolution of altruism

The idea that altruism might evolve has proven to be a contentious one. Many believe that, since evolution works at the level of individual reproductive fitness, serious other-directed behavior could not evolve, for any cost in individual reproductive fitness will cause the altruist to become gradually extinct. Still, we do observe behavior in humans and animals that seems to benefit the group more than the individual, and that has to be explained within the context of evolution. One might argue that evolution creates egoists but that does not necessarily mean that they are not ethical egoists: alternatively, one might argue that evolution works at the level of groups as well as individuals and therefore that group-directed behavior would be selected for as well.[8] Let's take a look at a few examples.

[7] **7** D. Dennett.Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon and Schuster, New York, 1995.

[8] E. Sober and D. Wilson.Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Boston, 1998.

Sugarscape

The first agent-based ALife model that attempted to simulate something like human societies is the ‘‘Sugarscape’’ program.[9] This model attempted to show how societies emerge from simple interactions between agents. Such fundamental collective behaviors as group formation, cultural transmission, combat and trade are seen to ‘‘emerge’’ from the interaction of individual agents following simple local rules.[10]

The agents are very simple. They populate an uncomplicated world with one resource, ‘‘sugar,’’ which they need to find and eat. They can then ‘‘metabolize’’ this resource and use it to move around to find more food.

Agents are born onto the Sugarscape with a sense of vision, a metabolism, a certain speed of movement and other genetic attributes. Their movement is governed by a simple local rule: ‘‘look around as far as you can; find the spot with the most sugar; go there and eat the sugar.’’[11] Every time an agent moves, it burns sugar at an amount equal to its metabolic rate. Agents die if and when they burn up all their sugar.[12]

From these simple beginnings a number of interesting social behaviors emerge. Migration when sugar grows at different rates along the landscape, hibernation when resources are low, and the formation of ‘‘tribal’’ grouping when the landscape is changed to facilitate this behavior. When the tribes migrate to a frontier area, ‘‘the two tribes interact, engaging in combat and competing for cultural dominance, to produce complex social histories with violent expansionist phases, peaceful periods, and so on.’’[13]

Epstein and Axtell have extended this modeling technique and have used it for a number of experiments in economics and social science, including a model that simulates genocidal behavior.[14]

Interesting as these models are, an awful lot of this emergent behavior is in the eye of the beholder. The groups are defined only very roughly and there is only a small attempt to model any sort of psychology. Even with these very simple agents, though, it is possible to get something that looks like group interactions. Epstein and Axtell’s hope is to find general social rules that emerge with little or no thought out of the simple interaction of individuals. Sugarscape does provide a very compelling simulation of the evolution of simple groups that

---

[9] J.M. Epstein and R. Axtell.Growing Artificial Societies: Social Science from the Bottom Up. MIT Press, Cambridge, 1996.
[10] Ibid.
[11] Ibid.
[12] Ibid.
[13] Ibid.
[14] J. Rauch. Seeing Around Corners: The New Science of Artificial Societies.The Atlantic Monthly, April, 2002.

emerge without external design and has inspired further work. ALife models that show evidence of cooperation, group selection and altruism

Another model was created by Pepper and Smuts at the Santa Fe Institute,[15] attempts to explore explicitly the emergence of altruistic behavior. In their paper, they begin by reviewing how David Sloan Wilson and Elliot Sober,[16] were able to show mathematically that if the right conditions hold, then natural selection can operate at the level of the group selecting for altruistic traits in individuals that benefit the group at large even though they exact a cost in fitness to the individual.

Wilson and Sober advanced this idea in the face of the well-known counter arguments from theorists such as Richard Dawkins who claim to prove that group selection is a dead theory because even if one can theoretically dream up schemes in support of group selection, it is meaningless because in the real world, groups are not stable enough to count as a ‘‘vehicle of selection.’’[17] To challenge the prevailing view, Pepper and Smuts have made a model where the programmer does not explicitly define the group level, which allows them to ask:

(1) Can spatial heterogeneity alone generate the local fitness effects and population structure necessary to drive group selection?
(2) Does group selection require association among kin in order to be effective? [18]

Essentially, Pepper and Smuts' model is a standard ALife cellular automaton that simulates a world with regular food patches, and is populated by two types of agents. The first kind of agent eats any food it encounters while the other, an altruist, restricts its feeding at a certain cost in reproductive fitness to itself. The behavior of the altruistic agent allowed more agents to survive in the midst of limited resources clearly benefiting the group.

Their model confirmed that altruistic behavior is selected for in between-group selection, although it was selected against in within-group selection. They found that, with patchy food distribution, the cooperative trait spread more quickly. Additionally, they discovered that the groups did not have to be composed of spatially located kin, thus countering any claims that this was simply a model of kin selection theory.[19]

The main finding is that the model questions the prejudice against group selection one finds in mainstream theoretical biology based on the idea that group selection needs, ‘‘discrete and

---

[15] J.W. Pepper and B.B. Smuts. The Evolution of
Cooperation in an Ecological Context: An Agent-based Model. In T.A. Kohler and G.J. Gumerman, editors, Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes. Oxford University Press, New York, 2000.
[16] Supra, No. 8.
[17] R. Dawkins.The Selfish Gene. Oxford University Press, Oxford, 1989.
[18] Supra, No. 15.
[19] Ibid.

stable groups that might not be typical in nature.''[20] This opens the door a bit wider for more serious inquiry into the exact nature of group selection and its effects on altruistic behavior and it begins to eliminate the notion that there can be no ethics or morality in a world created by the forces of evolution.

I think that the agent-based approach is the correct one but that the technology that we have today is a bit too weak to simulate the evolution of morality in full. To achieve this goal will require software agents that have at least some cognitive abilities. The more cognitive abilities the agents have the more interesting and useful these models will become. Of course, we are still in the dark about the proper way to model many cognitive processes, so it is no wonder that the models so far have stuck with very simple agents. It is just a matter of time, though, for more advanced programs to develop and, in the meantime, there is still much work that can be done with the extending simple models.

An important project that is currently attempting to address these robust problems is the New and Emergent World Models through Individual, Evolutionary and Social Learning (NEW TIES). This is a very ambitious three year research program begun in 2004 and funded by the European Union, which seeks to develop a very large and powerful agent based model where the agents might evolve language, society, and some understanding of their own existence.[21] NEW TIES involves a number of European researches in computer science, social science, and economics, all from different universities and countries, so there is a great deal of potential to this model. While the researchers do not specifically address the evolution of morality in their proposal, there is little doubt that this model, or one built with the technology they are developing, will have direct application to the study of the evolution of morality. It is unfortunate that those developing this model do not yet see this opportunity but they do acknowledge this ethical impact of their work:

...[T]he project has the opportunity (and the responsibility) to inform the wider public of the possible implications of the creation of large scale communities of autonomous virtual agents. We shall do this through the popular science media, endeavoring to provide a fair and objective view of future scenarios, without encouraging scare mongering.[22]

So far they have done so with a number of articles and brief news items appearing in popular magazines such as PC Magazine, New Scientist, and Information Week.[23]These researchers are correct, as we evolve more and more complex software agents and societies of software agents, we must address the issue of the moral status of software agents and recognize their capacities as fellow moral agents. I will develop this a bit more in the last section of this paper.

---

[20] Ibid.

[21] For a full description of this project see Annex I – escription of work Prepared on 05/04/2004 FP6-502386: NEW TIES (http://www.cs.vu.nl/%7Egusz/newties/an- nex1-short.pdf).

[22] Ibid.

[23] See (http://www.new-ties.org)

One thing that all of these models do achieve is providing increasing evidence that it is not inconceivable that societies and groups might be able to emerge out of evolutionary processes. Evolution is not the enemy of ethics and morality and, through the use of more advanced models; this will become more and more apparent.

So we have two ways in which soft ALife models may have ethical impacts. First, they may help model difficult scientific theories, like Darwinian evolution, and as such can play a role in resolving the debates regarding the ethics of teaching alternative theories to evolution. And the second is that they help make sense of difficult issues in theoretical evolutionary ethics.

Ethics and hard ALife

Some of the researchers in the field of ALife have made the claim that they believe under certain circumstances some software agents can be said to be alive.[24] While ''life'' is notoriously difficult to define, it is true that some software agents have many of the qualities such; as metabolism, self-reproduction, subject to evolutionary process, etc., that one might list as necessary life functions. So, while it is difficult to unequivocally state that ALife programs are alive, it is also difficult to state categorically that no ALife programs are alive. So we must take seriously, at least at the theoretical level, the ethical implications of creating artificially living computational agents.

Certainly if we were to build computational systems that had cognitive abilities similar to that of humans we would have to extend them moral consideration in the same way we do for other humans. Of course we are nowhere near to creating such entities but what about more modest cognitive agents? At what point do they need to be considered as objects of moral deliberation? Matthew Elton of the University of Sterling has suggested that if one is willing to extend moral concern to animals, then one must also logically extend that concern to any other cognitive agent that is similar to an animal, which implies that if one is committed to protecting animals from violence then one should not wantonly destroy software agents in video games.[25] This probably goes a bit too far given the level of technology found in the videogames of today. Even the best AI programs in existence exhibit very modest cognitive skills and therefore have very little claim to moral consideration. Still, as these programs advance in abilities we will have to revisit this problem. What is more worrisome is that the level of violence simulated in some games may indeed raise ethical questions. Even if the software agent is not the equivalent of a human in cognitive skills, they are portrayed in some games fully rendered on the computer screen as humans who bleed suffer and die like humans. If a clear distinction is not drawn by the game one has to wonder at our ability to inflict such mayhem even if it is just pretend.

Given that our ALife programs are not the equivalent of humans or even animals, then can we conclude that Hard ALife software agents deserve no moral consideration? Perhaps, but one final claim might be raised when one considers Hard ALife projects as attempts to create new

[24] Supra, No. 2.
[25] Supra, No. 6.

artificial ecologies. If this is granted, then we should deploy the lessons of environmental ethics towards these potential new environments. While I would not suggest that even the best hard ALife projects, such as Tom Ray's ''Tierra'' project approaches the level of complexity of a real environment, it is a grey area and future programs may become more and more like the environments we talk about in environmental ethics.

Computer ethics versus machine ethics

In a recent paper Michael Anderson et al., argue that it is time to consider the role that computer technology plays in ethical decision making.[26] While they seem to gloss over the fact that the field of philosophy of technology has argued persuasively for a number of years that technology is the physical manifestation of philosophical values including ethical values. What they do correctly point out is that computer technology is increasingly becoming much more autonomous in its interactions with the world and as such will be increasingly called on to make ethical decisions which were once the exclusive realm of humanity.

Anderson et al., claim that the field of machine ethics is distinct from that of computer ethics because it is dealing with the ethical deliberation of machines towards humans and other machines, whereas computer ethics is concerned with the proper use of computers as that use effects other humans.

The main idea of the paper is that machines need to be built that can do their own moral reasoning. This is an fascinating argument due to the fact that there already exists many semi-autonomous systems such as telerobotic weapons systems, medical knowledge base systems etc., that are already greatly impacting ethical decision making in life and death situations. In addition to this Anderson et al., argue that many of our ethical systems such as Act Utilitarianism and Prima Fascia ethics require very advanced calculations of causes and effects as well as the potential contradictions implicit in any given behavior, such that calculating them all accurately and quickly moves well beyond the ability of human. Thus what happens is that we rely only on our best guess, which is often weighted too heavily towards our own self-interests. Computer systems can accurately crunch large numbers and do so judiciously as long as they are properly programmed to do so. So they recommend we build an AI moral reasoning knowledge base similar to those used in other disciplines such as medical or mechanical diagnostics.[27] In addition, working with an ethical reasoning knowledge base such as the one proposed would allow one to more accurately describe any given situation and realize in a more precise way who are all the various parties that would be effected by any contemplated action and what those effects might entail.

Anderson et al., have built two systems, Jeremy and W.D. that implement Act Utilitarianism and Prima Fascia moral reasoning, respectively. The results are preliminary and still have a number of problems. One I noticed was that the user is asked to determine the effects of the

[26] Ibid.
[27] Ibid.

action under consideration on all of the agents involved in the situation. This means that the goal of creating more accurate ethical reasoning is compromised since the user's preconceived notions of utility and effect are just reinforced by the computational process. This would result in the agent being able to more easily justify an immoral act based on the authority of the so called impartial ethics knowledge base.

This flaw needs to be addressed and it might be a fatal flaw for the foreseeable future. This is due to the fact that the machine would have to have a nuanced understanding of the situation at hand, in which case it would need to be a fully functioning AI the likes of which do not exist today and may not for some time. Even if sophisticated ethics knowledge bases do come online there is no guarantee that they would be as impartial as is required of this system since they would be agents in their own right and have interests of their own clouding their reason just like we do.

Machine Ethics does raise a very important point though. It will probably be the case that we will have to program our robust ALife software agents to ethically deliberate about their actions and how they might effect humans who interact with the system, long before we have to ethically deliberate about our interactions with the same software agents. This is due to the fact that these systems will be more and more autonomous and as such direct human control will become less and less of a possibility so as designers of these programs we have to make sure they can limit for themselves the amount of harm that we know self-reproducing programs can do.

Extending the concept of computation and the emergence of wet ALife

Up until recently the field of ALife has focused on exploring issues in theoretical biology through the use of computer simulations and/or robotic models. Even though the arguments presented to claim that these systems meet the minimum criteria for life are subtle and often interesting, if one believes in some sort of primacy of biological matter, then these arguments may seem to be easily dismissed.

This argument is going to become more difficult to maintain if the field of Wet ALife succeeds. Some ALife researchers are taking their theories off the computer and instead manipulating the constituent pieces of cells to attempt to synthesize very primitive artificial living cells. [28]

This begins to move the ethical concerns of ALife out of computer ethics and into bio-ethics. But with the emergence of bio-computational devices it is inevitable that these heretofore separate ethical disciplines will have to merge as their interests begin to intersect.

[28] J. Szostak, D. Bartel, P. Luisi. Synthesizing life.Nature, 409: 383–390, 2001; S. Rasmussen, L. Chen, D. Deamer, D. Krakauer, N. Packard, P. Stadler, M. Bedau. Transitions from nonliving and living matter.Science, 303: 963–965, 2004.

So far there has not been any greatly significant results in this area of research, but if these bio-computational entities are successfully created, then they will be much harder to dismiss then the purely software simulations that have been created so far. When and if this happens it is likely that many traditional thinkers will find the success of wet ALife and anathema and the ethics of doing this research highly questionable much in the same way they object to stem cell research.

The anthropologist Stephan Helmreich has traced the impacts hard and soft ALife has already had with religion. In his book; Silicon Second Nature, 1998, he argues that ALife can serve for its practitioners as a kind of surrogate for traditional religion as well as provide strength to the sociobiological interpretation of human nature.[29] In the political landscape of today where traditional religious beliefs are on the upsurge worldwide, this alone will place ALife in the path of a great deal of popular criticism.

The role of philosophers in this debate

In 1994 Daniel Dennett wrote:

> There are two likely paths for philosophers to follow in their encounters with Artificial Life: They can see it as a new way of doing philosophy, or simply as a new object worthy of philosophical attention using traditional methods. ...I urge Philosophers to take the leap and consider the first to be more important and promising.[30]

So far only a few philosophers have taken Dennett up on his challenge, regardless of the fact that the field of ALife has been very accommodating to philosophers and philosophical issues. With Mark Bedau as a very notable exception, few professional philosophers regularly make use of ALife models. Perhaps it is just the philosopher's distrust in, and lack of facility with, computer models that has been driving this phenomenon.

Unlike hard or soft ALife, wet ALife is likely to be an even more unlikely tool for philosophers to employ directly. I agree that soft ALife is a potentially powerful tool in the construction of philosophical thought experiments. Mark Bedau writes that, ''...artificial life's computational methodology is a direct and natural extension of philosophy's traditional methodology of a priori thought experiments.''[31]The thought experiments found in ethics and morality can be quite inconclusive, and exploring them in terms of computational models will help expose hidden, or unworkable, assumptions contained in the main ethical theories found in philosophy today.

---

[29] S. Helmreich.Silicon Second Nature: Culturing Artificial Life in a Digital World. University of California Press, Berkeley, 1998.

[30] D. Dennett. Artificial Life as Philosophy.Artificial Life, 1(3): 291–292, 1994.

[31] M.A. Bedau. Artificial Life. In L. Floridi, editor, Philosophy of Computing and Information. Blackwell Publishers, 2003.

But I also believe Dennett's dismissal of indirect philosophical analysis of ALife is too hasty. There is quite a bit of philosophical analysis of the social impacts of ALife that is left to be done.

One impact that I am interested in is why society in general feels uncomforted by scientific explanations of life such as those offered by ALife. It should come as no surprise to ALife researchers that their research programs may come under fire from more traditional belief systems. If ALife is a correct hypothesis, we should see these opposing social movements as perhaps two species of competing memes. If we do so, it is obvious that they will try to out reproduce one another in a competition for space in the conceptual landscape of human culture. What will be the likely outcome of this struggle?

An uncomfortable possibility, suggested by the work of the theoretical biologist David Sloan Wilson, is that traditional beliefs, though factually false, are more group functionally adaptive than scientific belief systems such as ALife.[32] So over time the traditional beliefs will prove to be more fit and will drive the more scientific belief systems to extinction. This is a chilling idea; the accuracy of this intuition might possibility be studied using an ALife model similar to the NEW TIES model mentioned above.

Taking artificial moral agents seriously

In this paper I have shown that whether we are looking at hard or soft ALife, situated in vitro or in silico, this technology challenges our traditional notions of morality and moral agents. Morality is no longer, and may never have been, the exclusive claim of human agents. Certainly, autonomous ALife agents will become embroiled in moral situations, either through creating harms to other human or nonhuman agents or in suffering harm themselves. Therefore they must be considered moral agents.

Saying this runs afoul of a number of the traditional assumptions often made by moral philosophy, which suggests that a moral agent must have free will, intentional mental states, and is able to hold responsibility for its actions. Clearly, ALife agents have none of these things, so how can we sensibly call them moral agents?

Luciano Floridi and J.W. Sanders, of the Information Ethics Group at the University of Oxford argue that we need a ''mindless morality'' that would enable us to bypass these tricky traditional qualities of moral agents that, undeniably, have proven problematic to ascribe even to humans.[33]They argue that we can see artificial entities as agents when we properly set a certain level of abstraction where we can see the agent's actions as being interactive with their surroundings through state changes that are yet somewhat autonomous from their environment and also adaptable to new surroundings. When these autonomous interactions

[32] D.S. Wilson. Language as a Community of Interacting Belife Systems: A Case Study Involving Conduct To- wards Self ad Others.Biology and Philosophy, 10(1, January): pp. 77–97, 1995.
[33] L. Floridi and J.W. Sanders. On the Morality of Artificial Agents.Minds and Machines, 14.3: 349–379, 2004.

pass a threshold of tolerance and cause harm we can logically ascribe a negative moral value to them, likewise the agents can hold a certain appropriate level of moral consideration themselves, in much the same way that one may argue for the moral status of animals, environments, or even legal entities such as corporations.[34]

Floridi and Sanders provide many examples to back up their argument but one that is very convincing is the claim that programs written in such a way that their source code is closed must be treated as moral agents in a way that open source programs need not. Since a program that closes off the user from knowing its inner workings will appear, at the level of abstraction of the user, to be interactive and adaptive, and completely autonomous, since the user is unable to examine the program that determines its behavior, then the program is a moral agent that can harm the user if, for instance, it spies on the web viewing habits of its user, or some other breach of confidence. But at the level of abstraction of an open source program, it is no longer a moral agent since the user has complete scrutiny of the source code and is therefore more personally responsible for what the program does.[35]

This example works well for the purposes of this paper as it provides a very strong case for ascribing moral status to ALife agents. ALife agents are typically created using a genetic algorithm that tests each agent against some fitness function, allowing those that perform the best to reproduce in the next generation, and deleting those that do not. A typical ALife program may go thorough many thousands of generations, creating new agents each time without the aid of a human programmer. Given complex agents competing for survival in a complex environment, this process will result in nearly incomprehensible code more impenetrable than the best protected commercially produced code of today. This would mean that ALife agents could be seen, at certain levels of abstraction, as moral agents and in turn deserving of moral consideration themselves. Meaning that they may soon evolve to have a moral status similar to what we might ascribe to animals today. Floridi and Sanders' work stands as an important example of the role of philosophy in understanding and clarifying the moral implications of ALife technology.

Conclusion

ALife is a technology that is challenging to traditional ethics and morality on a number of levels. I have argued that regardless of whether you think ALife is simply a modeling technique or a way to synthesize novel living agents, either way, there are significant ethical challenges that this technology forces us to confront.

[34] Ibid
[35] Ibid

As a modeling technique, it allows us to make sense of the role morality may play in the evolution of societies. In addition, ALife is a powerful way for philosophers to extend traditional thought experiments regarding ethics and morality.

If ALife is capable of synthesizing life either in vitroor in silico, then there will be significant ethical implications given the potential impacts of these entities in their respective environments. But as we have seen, we do not have to envision these seemingly science fiction extremes, it is absolutely certain that ALife programs can be seen as non-living artificial agents that can exhibit artificial morality and are indeed real moral agents.

References

M. Anderson, S.L. Anderson and C. Armen. Towards Machine Ethics. In Proceedings of AAAI Workshop on Agent Organizations: Theory and Practice, San Jose, CA, July, 2004.

M.A. Bedeau. Philosophical Aspects of Artificial Life. In Towards a Practice of Autonomous Systems. Proceedings of the first European Conference on Artificial Life, pp. 494–503, MIT Press, Cambridge, 1992.

M.A. Bedau. Artificial Life. In L. Floridi, editor, Philosophy of Computing and Information. Blackwell Publishers, 2003.

R. Dawkins, The Selfish Gene. Oxford University Press, Oxford, 1989.

R. Dawkins. The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design. W.W. Norton and Company, Inc., 1996.

D. Dennett. Artificial Life as Philosophy. Artificial Life, 1(3): 291–292, 1994.

D. Dennett,Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon and Schuster, New York, 1995.

M. Elton. Should Vegetarians Play Video Games? Philosophical Papers, 2000.

C. Emmeche. Life as an Abstract Phenomenon: Is Artificial Life Possible? In Towards a Practice of Autonomous Systems. Proceedings of the first European Conference on Artificial Life, pp. 466–474, MIT Press, Cambridge, 1992.

C. Emmeche, The Garden in the Machine, the Emerging Science of Artificial Life. Steven Sampson, translator, Princeton University Press, Princeton, 1994.

J.M. Epstein and R. Axtell, Growing Artificial Societies: Social Science from the Bottom Up. MIT Press, Cambridge, 1996.

L. Floridi and J.W. Sanders. On the Morality of Artificial Agents. Minds and Machines, 14.3: 349–379, 2004.

S. Helmreich, Silicon Second Nature: Culturing Artificial Life in a Digital World. University of California Press, Berkeley, 1998.

J. Horgan. From Complexity to Perplexity. Scientific American, June: 104, 1995.

J. Horgan. The End of Science. Addison-Wesley, Reading, 1996.

C.G. Langton. Artificial Life. In Chris Langton, editor, SFI Studies in the Sciences of Complexity, Proc. Vol. VI. Addison-Wesley, Redwood City, 1989.

S. Levy, Artificial Life: The Quest for a New Creation. Pantheon Books, New York, 1992.

R. Lewontin, Complexity: Life at the Edge of Chaos. Macmillan Publishing Company, New York, 1992.

E.T. Olson. The Ontological Basis of Strong Artificial Life. Artificial Life, 3(1): 29–39, 1997.

H.H. Pattee. Simulations, Realizations, and Theories of Life. In C. Langton, editor, Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, held September 1987, in Los Alamos, New Mexico, Vol. 6. Addison-Wesley, Redwood City, CA, 1989.

J.W. Pepper and B.B. Smuts. The Evolution of Cooperation in an Ecological Context: An Agent-Based Model. In T.A. Kohler and G.J. Gumerman, editors, Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes, Oxford University Press, New York, 2000.

J. Rauch. Seeing Around Corners: The New Science of Artificial Societies. The Atlantic Monthly, April, 2002.

S. Rasmussen. Aspects of Information, Life, Reality, and Physics. In C. Langton, C. Taylor, J.D. Farmer, S. Rasmussen, editors, Artificial Life II: The Proceedings of the Workshop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10, pp. 767–774. Addison-Wesley, Redwood City, 1992.

S. Rasmussen, L. Chen, D. Deamer, D. Krakauer, N. Packard, P. Stadler and M. Bedau. Transitions From Nonliving and Living Matter.Science, 303: 963–965, 2004.

T. Ray. An Approach to the Synthesis of Life. In C. Langton, C. Taylor, J.D. Farmer S. Rasmussen, editors, Artificial Life II: The Proceedings of the Work-shop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10, pp. 767–774. Addison-Wesley, Redwood City, 1992.

E. Sober. Learning From Functionalism – Prospects for Strong Artificial Life. In C. Langton et al., editors, Artificial Life II: The Proceedings of the Workshop on Artificial Life, held February 1990 in Santa Fe, New Mexico, Vol. 10. Addison-Wesley, Redwood City, 1992.

E. Sober and D. Wilson, Unto Others: The Evolution and Psychology of Unselfish Behavior. Harvard University Press, Boston, 1998.

J. Szostak, D. Bartel and P. Luisi. Synthesizing Life. Nature, 409: 383–390, 2001.

J. Sullins. Knowing Life: Possible Solutions to the Practical Epistemological Limits in the Study of Artificial Life. **The Journal of Experimental and Theoretical Artificial Intelligence**, 12(1): 13–22, 2000.