

Artificial Intelligence and Artificial Life – should artificial systems have rights?

Susan Stuart

November 1994; slightly revised August 2003

Abstract

The defining features of intelligence, consciousness and life are not clearly prescribed. We cannot say of X that if, and only if, it has properties p , q and r will it also have intelligence, be alive, or be conscious. What we tend to look for when ascribing inner states to another system – I shall use the term ‘system’ to cover both organic and inorganic entities and I shall examine three different areas of this theory of ‘systems’: *intelligent systems*, *conscious systems* and *living systems* – is appropriate behaviour. Then on a basis of that behaviour it is prudent that we behave with the other system as though it has, for example, intelligence, even though we cannot be absolutely certain that it does [viz. [2]].

Work in artificial life (A-Life) suggests that it will only be a matter of a some short period of time, perhaps fifty to a hundred years, until we can create a living system. Whether such a system will be living in inverted commas, or quite naturally alive in an organic sense only it can know for sure, and it would seem wise to consider the rights we might wish to afford an artificially produced living system before it has the complex capabilities to decide for itself.

1 Intelligent systems: What counts as intelligence?

It has been suggested that the big difference between human systems and non-human systems is that the former are ‘intelligent’ in a way that the latter are not. Human systems are capable of all sorts of things that other animals are not, or so the story goes. For instance, they can create and interact in complex societies, and they have the ability to create symbols, assign meaning to them and use them in meaningful and mutually comprehensible ways. But don’t ants and bees interact in complex societies? It is true that such societies are a lot less complex than human society but if we consider the limited capabilities of an ant or a bee their society is an amazingly complex one for them.

So maybe the only thing that makes human systems truly different is their sophisticated use of language. Certainly, humans are capable of all sorts of complex thinking. Thinking about objects or events that are not directly perceivable, thinking about subjective events and emotions, and abstract thinking about mathematical entities or the problem of what constitutes justice. All such thinking makes use of a complex network of sophisticated languages and we would doubtlessly refer to that sort of thinking as *intelligent*. The kinds of thought we usually count as lacking intelligence are those which are ‘mindless’ like, for example, the actions of an ‘automaton’, in this sense a person who follows a goal that is not their own, and we call their behaviour mechanical; or that which is either wrong as in $2 + 2 = 7$ or irrational, for instance, acting without adequate forethought, perhaps because of our passions or emotions as in a *crime passionnel*.

Of these we might want to argue that if animals are following innate physiological drives then they are simply sophisticated automata. If, however, they do things that deviate from the satiation of instinctive needs, then we would have to admit that their

actions seem, at least, to be intelligent. For example, if my cat has made a connection between my being vertical and ambulatory and its being fed I would consider its ‘knocking-things-of-the-bedside-table’ behaviour as quite intelligent if its aim is to get me out of bed and on my feet.

But our definition of intelligence still seems vague. It depends upon our interpretation of thinking and we only know what other human systems think because they tell us and we understand what they are saying. Such reports are subjective in the sense that they are made by the subject about the way they perceive their world, and we ascribe to them, on the basis of this reported information, complex mental states like beliefs and desires. In other words we ascribe intelligence to them. Knowing what an animal thinks when it acts in a particular way is a lot less clear, for we can have no subjective report, and the behaviour we interpret as intelligent might just be conditioned.

But ascription of intelligence is not always made on a basis of linguistic behaviour, indeed it is more often based on our observation of non-linguistic behaviour, behaviour that we consider to be representative of perhaps high-level mental activity. And so our ascription of mental states to other systems is often made implicitly, that is, in the way that we interact with the system. For instance, we behave differently with another human being than we do with a cat, and with a thermostat and a video-recorder. Previous interactions have shown us that video-recorders require more understanding than thermostats because they are more complex and have a greater range of functions; and similarly with human beings and cats.

But have we got any nearer to a proper definition of intelligence? It would seem not, instead what we have is a sort of ‘as-though’ behaviour, which is to say that when we interact with other systems, depending on their actions and their specific context we behave ‘as-though’ they have corresponding mental states. This goes on until our interpretation is justified or proved otherwise in which case we adjust our behaviour to suit our new information. Which is to say, we use the information we have about something, be that of its innards or its external behaviour, as a gauge for judging how we should interact with it.

This is all very well when we can look inside something, as we can do with a thermostat or a video-recorder, but not so easy with other living systems. With thermostats and video-recorders we can look inside and see that they have a rather simple structure, a structure that implies that it is incapable of truly intelligent action. But we can’t just lop the head off a living system to see if it really has a brain, or some other sort of equivalent functioning system that would be complex enough to suggest to us that it is capable of high-level intelligence. Nor, come to that, do we have much experience of our own ‘wetware’, nor do we yet fully understand the relation between ‘wetware’ and intelligence. Indeed it is possible that we, or at least some of us, might be highly sophisticated androids like an advanced, or should I say ‘less determined’, version of Schwarzenegger’s *Terminator*!

Does our definition matter if we can’t know what’s on the inside and the system behaves as if it is intelligent? We still treat it as-though it is intelligent. Which is exactly what Alan Turing suggested we do, from which arose the Turing Test. The test is by now fairly well known, but basically it consists of a curtain with a computer and a person situated behind it. On the other side of the curtain is a person who directs questions to both. If the questioner is unable to distinguish between them on the basis of the answers they give the computer program will have passed the ‘Turing Test’. There will be no perceived difference between the two systems and the computer program will be considered to be intelligent.

Admittedly the test does not mean that the computer program is actually intelligent but it does offer us the same solution that we put to use every day. We can never tell if the system we are interacting with is really intelligent or is running a good simulation. Come to that if some system is running a good simulation of intelligent behaviour what would be the difference between the mere simulation and the real thing? Is there

any difference? Is it the same sort of difference between an authentic painting and a very good, but forged copy? Who can tell, and should we care? Certainly we would care very much if it was a painting for which we had paid a great deal of money, and I'm sure we would feel a grave loss of dignity if we had been duped by a machine into believing it was an intelligent human being.

There is an expressible difference between 'artificial' intelligence and artificially created intelligence. The first is intelligence which only masquerades as intelligence, which, if you like, runs a good simulation of intelligent behaviour; and the second is intelligence which is, as far as our knowledge can demonstrate, the same thing as *real* intelligence but which does not occur naturally in the system that is now seen to possess it. But there is a problem with this distinction: it can easily be expressed in language, but when it comes to recognising it in practice we encounter problems, for we do not yet know what to look for, either in the brain or the behaviour of the system. Indeed it is difficult to speculate about what it is we would need to know before we could make the distinction in practice. It might be that we would have to be able to recognise when a system is having first hand experience of something; that, crucially, there is something that *it is like* to be the experiencing system. Thus only when we can recognise the first hand nature of another system's experience could we say that the system is doing more than simulating experience.

2 Conscious systems: What counts as consciousness?

I would like to move on at this point to the question of consciousness. All sorts of different levels of consciousness are presumed to exist. Some simple examples of these are the dreaming states in sleep, day-dreaming states, and waking states. But there are others, like a difference between consciousness and self-consciousness. Human systems can describe in language their interaction with their world, so we know humans to be both reflective and reflexive, which means that they can see that their actions affect themselves. But if human systems didn't speak it would be difficult to recognise that there is such a difference between consciousness and self-consciousness, and this is a problem we encounter when dealing with non-human systems that have no language we can understand. We do readily attribute self-consciousness to human systems and to some other higher-order mammals, but would we ascribe self-consciousness to ants or bees? I'm sure that we would ascribe consciousness to them because they are aware of their environment and respond to the informational input they receive from their external stimuli, but we have a problem when it comes to the sort of self-consciousness we associate with high-level human systems. Can an ant, a bee or even a cat understand that their actions affect them? As yet, it's impossible to know. If they can understand that their actions affect them then only they will know that for sure, and the same, I think, can be said of machines. If we manage to make one conscious and even self-conscious then only it will be able to know this with any degree of certainty.

What we seem to have now is the same problem with consciousness that arose with intelligence. We attribute consciousness to those systems that we believe behave in a conscious way without taking that system to pieces to have a root around inside. What terrible consequences there could be if we had to examine the innards of everything we interact with on a conscious level! Not only would it be very time consuming it would also be very messy! The Turing Test has again to be sufficient otherwise there could never be an end to our doubt about any other system. It is a pragmatic approach to the problem and one that is adopted by Daniel Dennett in *The Intentional Stance* [2]. As Dennett often reminds us, it is only through such a stance that we can continue to make reliable judgements about the prospective action of the things with which we interact. And the justification for this theory seems to be that it happens to work, and work very reliably.

Trying to escape the problem by defining consciousness in terms of awareness is possibly mistaken because even non-human inorganic systems (machines) can be said to be aware, even though it is, at present, only a low level of awareness. Even systems with command lines have to be aware of their informational input if they are to respond in the correct way. But, at present this awareness is very low-level stuff and not one that we would consider to be on a par with our common conception of consciousness.

If consciousness is something we associate with learning and adapting to new situations, then it is an ability that is undoubtedly of use to non-human systems as well as human systems, especially when they have to escape from danger or think of new ways to find food or shelter. In the recent work of people like Marion Stamp Dawkins [1] and Keith Oatley & Jennifer Jenkins [5] an emphasis is placed on the emotional element of consciousness. Dawkins argues that emotion, and not intelligence, is required for consciousness, and Oatley proposes that consciousness has evolved to stop us in our tracks when a goal is suddenly frustrated. It allows us valuable thinking time when we need it most. In his conception of consciousness thought and emotion are not separable. We need the emotional reaction to signal a need for thought. Emotion seems to be a very important notion for the attribution of consciousness to ants or bees.

So what happens when we have to deal with inorganic systems? Would we, for instance, want to say that the tracking systems of guided missiles possess consciousness? It would seem that under the simple criterion of ‘awareness of their environment’ we might tentatively say ‘yes’, but if we are to demand an ‘evolved emotional response to their environment’ we have to respond with a definite ‘no’. In general I think we can say that we are only really happy to attribute consciousness to those systems that we consider to be alive, to have certain evolved characteristics. Nonetheless, when someone falls into a coma we do not say that they are no longer alive, so what counts as being alive?

3 Living systems: What constitutes life?

Again a definitional problem arises. What characteristics do we consider to be necessary for the attribution of life to something? What makes us believe that some things are alive and others are not? Is consciousness one of those characteristics? It would seem not for surely we consider trees, flowers and bacteria to be alive. They act in regular ways that facilitate their survival but they are not considered to have consciousness.

A tongue-in-cheek answer to the question “what counts as being alive” has been derived from the work in the eighteenth century of Jacques de Vaucanson. It has been called the duck test and goes like this: if something looks like a duck and quacks like a duck, then it belongs in the class labelled ‘ducks’. It is, of course, a subjective test but then so is the Turing Test and so is the basis of our interaction with other systems. It is the relatively new field of artificial life that is at present attempting to find an answer to the test, one that is less subjective and does not admit of blood-letting for us to have any degree of certainty in our interactions. To make artificial life’s quest more intelligible I would like to give a couple of examples from recent work in the area. The first is a quotation from James Doyne Farmer:

Within fifty to a hundred years a new class of organisms is likely to emerge. These organisms will be artificial in the sense that they will originally be designed by humans. However, they will reproduce, and will evolve into something other than their original form; they will be “alive” under any reasonable definition of the word....The advent of artificial life will be the most significant historical event since the emergence of human beings.... [4, p.815]

This is an exhilarating paragraph. A feeling I am sure that is shared by Larry Yaeger who has created *PolyWorld*. *PolyWorld* is a study of a world inside a computer

whose inhabitants are made of mathematics! They have a digital DNA – I suppose in very much the same way as we have digital DNA that carries instructions for our personal characteristics in a coded form. In a way similar to the Darwinian survival of the fittest there are inhabitants that are fitter than others and it is those inhabitants that reproduce. From these reproductions a variety of creatures develop, adapt to and make use of the features of *PolyWorld*.

One description of the existence of such creatures is given by Steven Levy:

The creatures cruise silently, skimming the surface of their world with the elegance of ice skaters. They move at varying speeds, some with the variegated cadence of vacillation, others with what surely must be firm purpose. Their bodies - flecks of colors that resemble paper airplanes or pointed confetti - betray their needs. Green ones are hungry. Blue ones seek mates. Red ones want to fight.

They see. A... neural network bestows vision, and they can perceive the colors of their neighbors and something of the world around them. They know something about their own internal states and can sense fatigue. They learn. Experience teaches them what might make them feel better or what might relieve a pressing need.

They reproduce. Two of them will mate, their genes will merge, and the combination determines the characteristics of their offspring. Over a period of generations, the mechanics of natural selection assert themselves, and fitter creatures roam the landscape.

They die, and sometimes before their bodies decay, others of their ilk devour the corpses. In certain areas, at certain times, cannibal cults arise in which this behavior is the norm. The carcasses are nourishing, but not as much as food that can be serendipitously discovered on the terrain. [6]

This account is certainly florid, if not perhaps really quite emotive, but it is a still an account or description of a particular set of events that have actually taken place. It is not fiction. When each event is considered singly it is probably very simple. The mathematics that guides these creatures and from which they are made need only be very straightforward, but when the simple actions are put together, when they interact, the whole thing becomes quite complex and new patterns of behaviour emerge.

Artificial life suggests that human complex society is one such emergent property. It is one that has arisen from the way in which, over millions of years, we have used, implicitly, the information stored in our genetic codes and the information we receive and select from our interaction with our environment. And certainly the *PolyWorld* model is analogous with human societies where each individual follows simple rules, for example, blink when my eyes need moisture, find food when I'm hungry, don't step out in front of an oncoming bus and so on. When all individuals follow simple rules what develops is a complex interacting society which is quite different from the simple rules by themselves. This complexity of interaction is an emergent property and can also be seen in, for example, ant colonies and bee hives. Each ant or bee has a specific action to carry out but through the differences in their respective actions and the interactions that result, a complex colony or hive can emerge and thrive.

The term 'emergent properties' is derived from Gestalt psychology which states that "The whole is always greater than the sum of its parts". Gestalt psychologists wish to argue that properties emerge from a whole thing that do not emerge from the collection of its parts. An over-simplified example of this notion is to imagine a washing machine as bits and pieces disassembled in a pile on the floor; they can do nothing except rust. However, when those pieces are put together in an appropriate way we get the emergent property of the bits and pieces now being able to wash clothes. There are two things that are important here. The first is that all the pieces have an important role to play even if it is only a very simple one like holding the drum in place, and the second is the fact that they are organised in a particular way to make the complex interactions possible.

It is true that, to some extent, the washing machine example is a bit of a cheat because we know what property it is that we want to emerge from the correct configuration of pieces, we want to wash clothes. Yet, in our washing machine example a metaphysical point is being made. For instance, the washing is an extra property that has been derived from its constituent parts, yet nowhere could you look and say “Ah, that’s the bit that does the washing”. The washing emerges from the correct configuration of the totality of pieces.

With living systems it is not so easy, we don’t know the emergent behaviour we are aiming for. What could emerge from complex living systems is something that might possibly be irreducible to physical facts and relations. Again a metaphysical point is being made. As yet, we don’t know what information underlies the pieces that constitute a human being, let alone what sort of properties might have emerged or even still be emerging. But we are getting there, and more swiftly than we might imagine.

One possibility might be that the mind and its complexities, like language and emotions, are complex emergent properties that would have been neither predictable nor, as yet, explainable. Only with time and more information will it be possible for us to identify and explain such properties. But if consciousness is a property that has emerged out of the complexity of the organic system’s interaction with the world then it might be possible for consciousness, perhaps of a somewhat different nature, to emerge from a suitably complex inorganic system that has been supplied with the necessary information and the best configuration of that information. Indeed the onus now seems to be on those who disagree with what I am proposing, to explain why consciousness could not be an emergent property of organic systems so far, and a possible emergent property of sufficiently complex inorganic systems.

Here I want to introduce a third example, that of Craig Reynolds’ *Boids* [7]. Reynolds was greatly concerned about the sort of flocking behaviour exhibited by birds and fish. Their behaviour seems quite complex even though they are all acting on their own localised perceptions and not some sort of group intention. Reynolds devised three simple rules from which such complex behaviour might emerge: (i) a clumping action that tells the boids that they should keep together in their group, (ii) an ability to match the velocity of the group, so that they can speed up, keep up with and slow down with the group, and (iii) a separating force that stops them bumping into each other. When Reynolds implemented these rules within a computer environment the boids did behave in a flocking manner. So from simple behavioural rules a complex series of actions had emerged. Even when obstacles in the form of columns were introduced the boids managed to fly around them, even splitting the group for a couple of moments so that they could all avoid knocking into the new obstacles. On one occasion one boid did knock in to a column, it fell to the ground, remained there momentarily and then ‘shook itself’ and moved off to catch up with the group!

It seems entirely plausible that we are mere information processing systems as John von Neumann first proposed in the 1950s. We do just what the boids do, though in a much more complex environment with a lot more information. Our cells interpret the information stored in our genetic codes, and from the moment we are born we set about interpreting information from the outside world. It is this informational input that makes us who we are and the specific nature of the informational input of each individual that makes each of us emerge with a personality that is different from all others. And it is this dependency on information that makes us so similar to the systems created by artificial life. They have an informational base in their mathematical structure and they process information from their external environment. Living systems in all kinds of environments have to adapt strategies for sustaining their life and it would seem that in the PolyWorld example this is exactly what the mathematical creatures are doing.

These ideas have been prevalent in science fiction in one form or another for many years but they are no longer taken from a realm of fiction which is easily dismissible.

They are now part of science fact. And if we create a truly living system, one that has an autonomous existence, even one that possesses consciousness, do we have the right to terminate its life or do we have to then give it the same rights that we would any other living system? If we can create a living system, and the conservative estimate of Doyne Farmer makes this a possibility well within the lifetime of many of us, we should think carefully about its status in our society.

In his work on nanotechnology Eric Drexler has arrived at very much the same conclusion: "... genuine AI will arrive. To leave it out of our expectations would be to live in a fantasy world. To expect AI is neither optimistic nor pessimistic: as always, the researcher's optimism is the technophobe's pessimism. If we do not prepare for their arrival social AI systems could prove a grave threat: consider the damage done by the merely human intelligence of terrorists and demagogues." [3, p.81]

Given that artificially intelligent systems and, of course, A-Life systems are capable of learning, self-adaptation and reproduction would it be wise to let them reproduce at will? Do we really need another competing life form on an already crowded planet? It is possible that we could eventually be over-run? It would be like an invading army that we had invited in! Systems such as these have an immense amount of information, and can interact readily with their - and our - environment, so would we permit them to vote? Surely they would be better informed than many people who now vote, and better able to predict long term consequences of their actions since they have both speed of processing and a huge capacity for storing information on their side. What would happen if they turn out to be our intellectual superiors, would we have the right to use thuggery to control them? Thuggery in the sense of manipulating their 'minds', curtailing their aspirations, making them wish only to serve us? This sort of action was described in Huxley's "Brave New World" and Orwell's "1984" and neither sort of control was in any way admirable.

These problems about the ascription of rights to other non-human inorganic living systems are now upon us, and we had better be prepared just in case the new systems prefer a society of their own, and by exercising their superior intellect they might decide that it is we who should leave!

Susan Stuart
Humanities Advanced Technology and Information Institute
University of Glasgow
G12 8QQ
UK

Email: s.stuart@philosophy.arts.gla.ac.uk
<http://www.gla.ac.uk/departments/philosophy/Personnel/susan/>

References

- [1] Marion Stamp Dawkins. *Through Our Eyes Only?* W H Freeman/Spektrum, Oxford, 1994.
- [2] Daniel C Dennett. *The Intentional Stance*. MIT Press, 1988.
- [3] K E Drexler. *Engines of Creation: the Coming Era of Nanotechnology*. Oxford University Press, 1992.
- [4] J D Farmer. Artificial life: The coming evolution. In *Artificial Life II*, volume 10, page 815. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, 1992.
- [5] Jennifer Jenkins and Keith Oatley. *Understanding Emotions*. Basil Blackwell, 1996.
- [6] S Levy. *Artificial Life: A Quest for a New Creation*. Jonathan Cape, 1992.
- [7] C Reynolds. Flocks, herds and schools: A distributed behavioral model. *Computer Graphics*, 21:25, July 1987.

