Ethics Inf Technol (2010) 12:223–234
DOI 10.1007/s10676-010-9218-6

ORIGINAL PAPER

Implications and consequences of robots with biological brains

Kevin Warwick

Published online: 12 February 2010

Abstract In this paper a look is taken at the relatively new area of culturing neural tissue and embodying it in a mobile robot platform—essentially giving a robot a biological brain. Present technology and practice is discussed. New trends and the potential effects of and in this area are also indicated. This has a potential major impact with regard to society and ethical issues and hence some initial observations are made. Some initial issues are also considered with regard to the potential consciousness of such a brain.

Keywords Robot ethics · Robot rights · Human ethics · Consciousness · Robotics · Autonomy

Introduction

In the world of fiction a number of stories have attempted to tackle the complex issue of re-embodiment of a brain of one type or another, along with the ramifications that such an event brings about. The body being a means for the brain to receive sensory inputs and for outputs from the brain to affect the external world, possibly through movement. In effect the body is a bidirectional interface between the brain and the world.

One good example is to be found in Kafka's Metamorphosis (Kafka 1972). In this tale the hero, a human, wakes one morning to find that his body has turned into that of a bug. It is indeed interesting to follow how he has to learn to walk again now that he has many more legs to contend with. However, the story revolves mainly around how he is treated, in his new guise, by his friends and family. His family in fact appears to remain remarkably calm in the circumstances. But the hero of the story (Gregor) has major problems in communicating with them and seems to lose his taste for traditional foods. Disappointingly Kafka avoided the more difficult issue of the change in sensory signals that would no doubt have occurred and Gregor's brain seemed to emerge in very much its original form, surprisingly untraumatised.

The topic was also viewed by Moravec in his "Mind Children" (Moravec 1990) in a more modern setting. Here Moravec considered the possibility of copying, cell for cell, a human brain from its biological, carbon original form into a silicon, computer version. The latter entity then has the enviable opportunity to reside within a robot body, with all its advantages such as easy replacement limbs. Whilst Moravec did revel at the possibility of this new version living forever, again he appeared to overlook the trauma that might be caused when a brain suddenly realizes that all sensory inputs are different and movement is altered beyond all recognition.

Indeed such trauma may well have been experienced by animals involved in the past in head transplant experimentation, especially monkeys and dogs (White et al. 1963). There are certainly important differences though, one feels, between a brain, encased in a head, which is given a different body—with their original head remaining intact, and a brain being given an entirely new body, head included. In comparison, Kafka's Gregor expounds a seeming casualness to this eventuality. But recent fictional storylines have focused more on the dangers apparent due to the new resultant individual quite simply being different!

K. Warwick (✉)
School of Systems Engineering, University of Reading, Whiteknights, Reading RG6 6AY, UK
e-mail: k.warwick@reading.ac.uk

Clearly the topic of mixing and matching brains and bodies has provoked interest across cultures. For each of the stories mentioned, many more exist investigating previously unexplored concepts with sometimes horrific consequences. Mary Shelley's Frankenstein (Shelley 1831) is a prime example. She even explored what the monster thought and the problems he faced. But then, when restored to life in another human body, presumably senses and motor skills are not going to be too far removed from their originals. By comparison Kafka's Gregor really did draw the short straw. Frankenstein's monster though did produce a kneejerk reaction from acquaintances—many being scared and defensive against the new life form that they now witnessed—perhaps such is a natural human reaction. Is it, and should it be, the same case for a newly observed robot life form?

It is one thing to merely speculate and develop a storyline in a scientific vacuum, it is quite another to investigate what is actually going on when science does a catching up exercise. As we have seen, mixing human neural processing, through implants, with machine intelligence and the internet gives the upgraded human extended abilities and opens up many ethical and moral questions (Warwick 2003). It is now also quite possible to grow a biological brain and allow it to develop within a robot body (Marks 2008; Warwick et al. 2010). The end result is a robot with a biological brain.

Whilst the size and power of such a brain is trivial in comparison with that of a human brain then the issues are arguably limited. But when brain power is comparable then the problem clearly needs to be confronted. It is worth noting that the term 'brain' is employed here to describe the centre of the nervous system—which is responsible for the generation of behaviors and for extracting information from the environment.

In the section which follows, in order to familiarize the reader, the technology and processes involved are briefly described. Subsequently advances in the field are discussed along with future potential developments. The resultant implications of such ongoing technological changes and opportunities are then considered. It is apparent that this is just making a start on the subject. However, when considering the ethical implications of robots in general, merely to look at those which have a computer/machine brain would only be investigating part of the issue—robots with biological brains and robots with hybrid brains present considerable problems which need to be attended to.

Technology involved

The intelligent controlling mechanism of a typical mobile robot is usually a computer or microprocessor system and hence much of the initial work considering the future ethics and rights of robots has apparently focused only on this specific sub-class of intelligent robots (e.g. Arkin 2009). Research is however, now ongoing in which biological neuronal networks are being cultured and trained to act as the brain of a physical, real world robot—either completely replacing or operating in tandem with a computer system.

From a medical viewpoint, studying such neuronal systems can help us to understand biological neural structures in general and it is to be hoped that it may lead to basic insights into problems such as Alzheimer's and Parkinson's Disease. Other linked research meanwhile is aimed at assessing the learning capacity of such neuronal networks (Xydas et al. 2008). To do this a hybrid system has been created incorporating control of a mobile wheeled robot solely by a culture of neurons—a biological brain.

A brain, the human version in particular, can be said to be a complex computational platform (Lloyd 1991). It rapidly processes a plethora of information, is adaptable to noise and is tolerant to faults. Recently however, significant progress has been made in the practical integration of biological neurons and electronic components by culturing tens of thousands of brain cells in vitro (Bakkum et al. 2003). These technologies blur the distinction between synthesized brains and those which can be regarded as being achieved through a normal biological route.

The brains are created by dissociating/separating the neurons found in cortical tissue using enzymes and then culturing them in an incubator, providing suitable environmental conditions and nutrients. In order to connect a brain with its robot body, the base of the incubator is composed of an array of multiple electrodes (a multi electrode array—MEA) providing an electrical interface to the neuronal culture (Thomas et al. 1972).

Once spread out on the array and fed, the neurons in such cultures spontaneously begin to grow and shoot branches. Even without any external stimulation, they begin to re-connect with nearby neurons and commence both chemical and electrical communication. This propensity to spontaneously connect and communicate demonstrates an innate tendency to network. The neuronal cultures themselves form a layer over the electrode array on the base of the chamber making them accessible to both physical and chemical manipulation (Potter et al. 2001).

The Multi Electrode Array enables output voltages from the brain to be monitored from each of the electrodes, allowing the detection of the action potential firing of neurons near to each electrode as voltage spikes representative of neural charge transfer. It is then possible to separate the firing of multiple individual neurons, or small groups, from a single electrode (Lewicki 1998).

With multiple electrodes an external picture of the neuronal activity of the brain can be pieced together. It is also possible to electrically stimulate any of the multiple electrodes in order to induce neural activity. The Multi Electrode Array therefore forms a functional and non-destructive bi-directional interface with the cultured neurons.

Effectively, via certain electrodes, the culture can be stimulated and via other electrodes the culture's response can be measured.

As part of the research program, a disembodied cell culture can be provided with embodiment by placing it in a robot body. This is meaningful and sensible since a dissociated cell culture growing in isolation and receiving no sensory input is unlikely to develop much useful operation because sensory input significantly affects neuronal connectivity and is involved in the development of meaningful relationships necessary for useful processing. Hence the biological culture can be given a robot body such that signals from the robot's sensors stimulate the brain whilst output signals from the brain are employed to drive the motors of the robot.

Several different schemes have thus far been constructed in order to investigate the ability of such systems. Notably, Shkolnik created a scheme to embody a culture within a simulated robot (Shkolnik 2003). Two channels of a Multi Electrode Array, on which a culture was growing, were selected for stimulation and a signal consisting of a ±600 mV, 400 μs biphasic pulse (that is a pulse which is first positive then negative) was delivered at varying intervals. The concept of information coding was formed by testing the effect of electrically inducing neuronal excitation with a given time delay between two stimulus probes. This technique gave rise to a response curve which forms the basis for deciding the simulated robot's direction of movement using simple commands (forward, backward, left and right).

In a later well publicized experiment, DeMarse and Dockendorf also investigated the possibilities apparent with cultured networks by introducing the idea of implementing the results in a "real-life" problem, namely that of controlling a simulated aircraft's flight path (e.g. altitude and roll adjustments; DeMarse and Dockendorf 2005).

Present day ongoing embodiment

Present day ongoing research involves the removal of the neural cortex from the fetus of a rat. Enzymes are then applied to disconnect the neurons from each other. A thin layer of these disassociated neurons is subsequently smoothed out onto a Multi Electrode Array which sits in a nutrient rich bath. Every couple of days the bath must be refreshed in order to both provide a food source for the culture and to flush away waste material.

As has been already stated, as soon as they have been laid out on the array the neurons start to reconnect. Initially these can be regarded as mere projections, but subsequently they form into axons and dendrites, making connections between neighboring neurons. By the time the culture is only 1 week old, electrical activity can be witnessed to appear relatively structured and pattern forming in what is, by that time, a very densely connected matrix of axons and dendrites.

The Multi Electrode Array presently employed by my own research team consists of a glass specimen chamber lined with an 8 × 8 array of electrodes as shown in Fig. 1. The array measures 49 mm × 49 mm × 1 mm and its electrodes provide a bidirectional link between the culture and the rest of the system.

Thus far we have successfully created a modular closed loop system between a (physical) mobile robotic platform and a cultured neuronal network using the Multi Electrode Array method, allowing for bidirectional communication between the culture and the robot. It is estimated that each culture employed consists of ~100,000 neurons. The electrochemical activity of the culture is used as motor input to drive the robot's wheels and the robot's (ultrasonic) sensor readings are (proportionally) converted into stimulation signals received by the culture as sensory input, effectively closing the loop and giving the culture a body.

A Miabot robot has been selected as the physical platform. This exhibits very accurate motor encoder precision and speed. Hence the signals passing to and from the culture have an immediate and accurate real world physical meaning. Figure 2 shows the physical (Miabot) robot

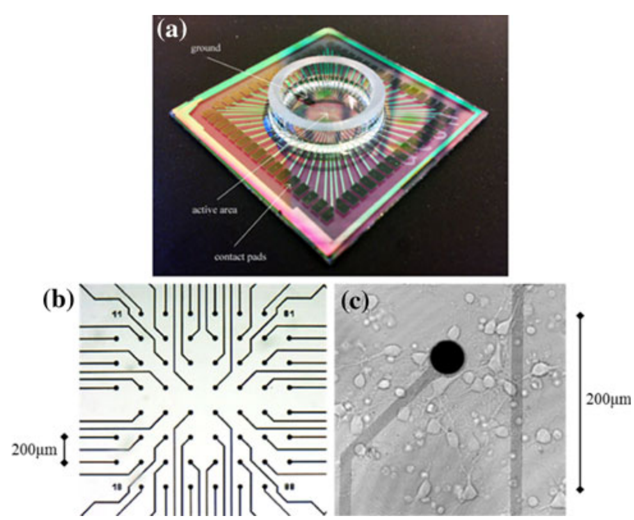

Fig. 1 **a** A Multi Electrode Array showing the 30 μm diameter electrodes, **b** electrode in the centre of the MEA seen under an optical microscope, **c** ×40 magnification, showing neuronal cells in close proximity with visible extensions and inter-connections

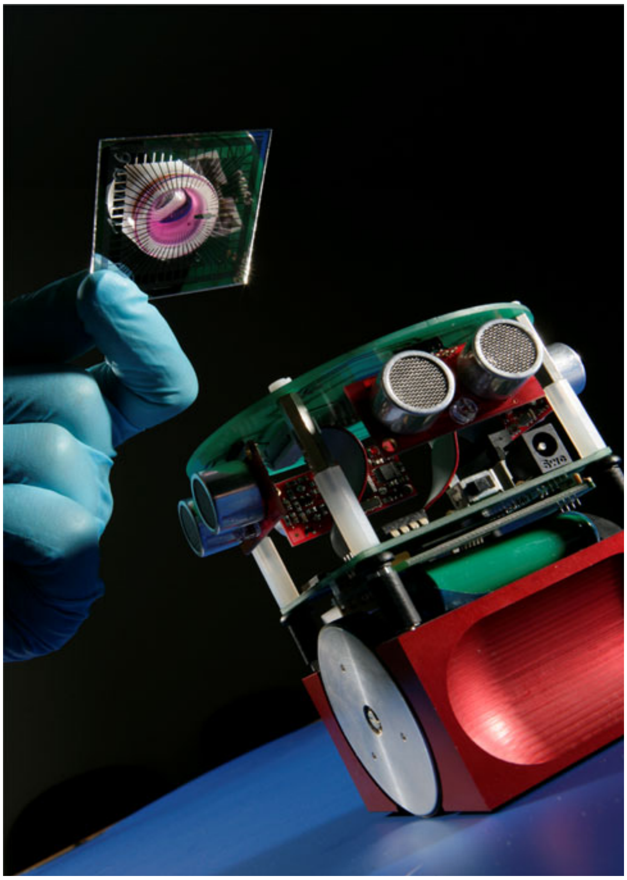

Fig. 2 Multi electrode array with culture, close to Miabot robot

employed along with an adjacent culture on a Multi Electrode Array—body and brain together.

The Miabot robot is wirelessly controlled from the culture in the incubator via a Bluetooth connection. Communication and control is performed through custom server code. A router/server computer has direct control of the Multi Electrode Array recording and stimulating. software. The server sends motor commands to the robot and feeds back sensory input to the culture.

It is worth pointing out here that in general a brain and its physical embodiment do not have to be within a confined body in one place. As long as a suitable neurological connection (effectively an extended nervous system) is in place, so a brain can be in one physical place and some (or all) of its body can be elsewhere. In this case the cultured brain exists in an incubator whereas its body is in an adjacent corral. Experiments with regard to such remote embodiments for humans have successfully placed robot body parts on different continents (Warwick et al. 2003, 2004). It is worth noting that this situation is somewhat different from the Brain in a Vat concept which has been studied extensively in philosophical literature (e.g. Brueckner 1986), as in the case considered here a strong interconnection exists between the brain and the outside world, even though there may be considerable distance and hence a lengthened time delay, between the two.

Experimentation

We have already conducted a series of experiments utilising a live culture. Initially, an appropriate neuronal pathway within the culture was identified and suitable stimulus electrodes and response/motor electrodes were chosen. The selection was made based on the criteria that the response electrodes show minimal spontaneous activity in general but respond robustly and reasonably repetitively to the stimuli (a positive-first biphasic waveform; 600 mV; 100 µs each phase) delivered via the stimulating electrodes. These spontaneous events were deemed 'meaningful' only in the case when the delay between stimulation and response was less than 100 ms. In other words, an event was a strong indicator that the electric stimulation on one electrode caused a neural response on the recording electrode (Warwick et al. 2010).

The overall task the robot had to achieve was merely moving forwards in a corral and not bumping in to any object, for example a wall. To this end, in the experiments conducted, the robot followed a forward path within its confines until it reached a wall, at which point the front sonar value dropped below a set threshold value triggering a stimulation/sensory pulse to be applied to the culture. If the responding electrode registered activity following the pulse, the robot turned in order to avoid the wall.

In its early life the robot sometimes responded "correctly" by turning away from the wall, although it also bumped into the wall on numerous occasions. The robot also sometimes turned spontaneously when activity was registered on the response electrode even without a stimulus pulse being applied. The main results to be highlighted though were the chain of events: Wall Detection–Stimulation–Response.

One point of interest was the maximum speed at which the closed loop system could respond, which was clearly dependant on the "thinking" time delay in the response of the culture. By itself this presents an interesting study into investigating the response times of different cultures under different conditions and how they are affected by external influences such as electrical fields and chemical stimulants (e.g. Cannabis).

As a follow up closed loop experiment the robot's individual (right and left separately) wheel speeds were controlled from the two chosen response/motor electrodes. Meanwhile received sonar information was used to directly control (proportionally) the stimulating frequency of the two sensory electrodes.

Run-times have thus far generally only been executed for ~ 1 h at a time, however, the robot's corral is being fitted with a special purpose powered floor which will subsequently allow for the possible study of a culture being embodied 24 h a day, 7 days a week over an extended period. It will then be of considerable interest whether or not the culture requires much in the way of down time (sleep equivalent), how quickly its performance improves and if its useful lifespan increases.

Presently a 'wall to stimulation' event corresponds to the 30 cm threshold being breached on the sensor such that a stimulating pulse is transmitted to the culture. Meanwhile a 'stimulation to response' event corresponds to a motor command signal, originating in the culture, being transmitted to the wheels of the robot to cause it to change direction. It follows that for the culture some of the 'stimulation to response' events will be in 'considered' response to a recent stimulus—termed meaningful, whereas other such events—termed spontaneous—will be either spurious or in 'considered' response to some thought in the culture, about which we are unaware.

Learning characteristics

Inherent/innate operating characteristics of the cultured neural network are taken as a start point to enable the physical robot body to respond in an appropriate fashion. The culture then operates over a period of time within the robot body in its corral area. This experimentation presently takes place once every day for an hour or so. Although learning has not, as yet, been a focus of the research, what has been witnessed is that neuronal structures that bring about a satisfactory action apparently tend to strengthen purely through the habitual process being performed. This is though mainly an anecdotal observation at this time, which is presently being formalized and quantified through further more extensive studies.

At first a stimulation-motor response feedback action occurs on some, but not all, occasions, and the action can be brought about sometimes without any sensory signal being applied—because the culture 'feels like it'. After habitually carrying out the required action for some time, the neural pathways that bring this about appear to be strengthened—referred to as Hebbian learning (Hebb 1949). As a result of this learning, "appropriate" actions therefore gradually become more likely to occur and spurious, unprovoked decisions to suddenly turn become less likely.

Research is now ongoing to use other learning methods to quicken the performance upgrade—reinforcement learning being one example. One major problem with this is deciding what exactly the culture regards as a reward and what as a punishment.

Comments on the culturing methodology

The culture preparation techniques employed are constantly being refined and have lead to successful and stable cultures that exhibit both spontaneous and induced spiking/bursting activity.

A stable robotic infrastructure has also been set up, tested and is in place for future machine learning and culture behaviour experiments. The current rate of progress could also lead to projects investigating culture-mediated control of a wide array of additional robotic devices, such as robotic arms/grippers, mobile robot swarms and multi-legged walkers.

There are a number of ways in which the current system will be expanded in the future. Indeed the Miabot robot is presently being extended to include additional sensory devices such as audio input, further sonar arrays, mobile cameras and other range-finding hardware such as an on-board infra red sensor. A considerable limitation is however, the battery power supply of an otherwise autonomous robot.

A main present consideration is therefore, as previously mentioned, the inclusion of a powered-floor for the robot's corral, to provide the robot with relative autonomy for a longer period of time while the suggested machine learning techniques are applied and the culture's behavioural responses are monitored.

The current hardcoded mapping between the robot goals and the culture input/output relationships will be extended to Machine Learning techniques which will ultimately reduce or completely eliminate the need for an apriori mapping choice. Reinforcement Learning techniques will then be applied to various mobile robot tasks such as wall following and maze navigation.

One key aspect of the research is a detailed study of the cultured neural network in terms of its observed connectivity density and activity in response to external stimuli. This behavioural evaluation is likely to provide great insight into the workings of the neuronal network by comparing its performance relating the culture's learning capabilities in terms of its neural plasticity.

Observations

It is normal practice for several cultures to be started at the same time. A typical number may be 25 different cultures. By using the same Miabot robot body it is then possible to investigate similarities and differences between the cultures. Clearly each culture is unique in itself, has its own individual identity (in the sense of it being recognizable—see Lloyd 1991 for further discussion) and is dependant on the original neural layout in its case and its own subsequent growth and development.

With regard to robot performance such cultural differences can be realised in terms of a robot which performs with fewer mistakes, one that responds more quickly or slowly, one that does its own thing more often or perhaps responds only after several signals are received. In essence there can be a large number of observed differences in behaviour even with a relatively simple task to be performed.

When investigating the behavioural response of an animal it can be difficult to ascertain neural differences because the overall neural requirements of the animal are not particularly understood, indeed many can appear as meaningless to humans. The advantage with our robot system is that its behavioural repartee can be investigated directly in terms of neural development, even in response to the effect on the culture of small changes in the environment. What the culture is "thinking" to itself though is merely a matter of conjecture.

In its early life the culture exhibits bursting activity, wherein spontaneous electrical activity can be witnessed on all (or at least most) of the electrodes. It appears that this is all part of the culture's development, but at this time exact reasons for it are unclear. As the culture ages so such bursting diminishes.

Cultures can be kept alive for perhaps 2 years or even more. After about 3 months or so in present studies however, they become much less active and responsive and hence most research involves cultures aged between 1 week and 3 months. This period is though sufficient to consider culture development and neural pathway strengthening. Present lifetime expectancy is limited due to a cultures vulnerability to viruses and the need to establish rigorous environmental growth conditions. Ongoing research in this area, into which conditions are preferable and which are less favourable, is however, improving the life expectancy of cultures.

When connected in its robot body a culture exhibits regular neural pathway firings. A few of these can potentially be directly diagnosed as being something to do with stimulating sensory signals, however, the vast majority cannot be so classified. The nature of other connections and signalling can only be guessed at. Certainly some neurons adjacent to a stimulating electrode appear to play more of a role as targeted sensory input neurons. Meanwhile others adjacent to output electrodes appear to take on more of a role as motor neurons. Yet again there are other neurons that appear to play a routing, controlling activity. Such specialisation seems to arise naturally through the culture's development. But the exact nature and role of each of these neurons is of course mere speculation and will necessarily remain as anecdotal observation.

When positioned in its robot body it is easy to relate neural firings that link to one another in response to particular sensory stimulating signals and/or decisions taken by the culture for specific motor outputs. What is not so straightforward however, is explaining such types of firings when the culture is disembodied and merely sitting alone in the incubator. Such a case is relatively normal for the culture but is not (I stand to be corrected) experienced by a regular animal or human whose brain lives its entire life receiving sensory input and making motor output decisions—other than possibly when in a sleep and/or dream state. Within the incubator both bursting and structured neural firings can be witnessed in the culture. The question arises as to what these firings mean.

When observing the activity in a culture it is natural to speculate with a plethora of questions. When the culture is disembodied, does it dream? If not, what is it thinking about? What must it 'feel' like to be the culture? Do the firings relate to previously experienced sensory stimulation that it is reliving? Does a brain need external stimulating signals in order to subsequently make up stories in itself? If more stimulating signals are applied, of a different type, will such disembodied signals be more and/or different?

Some questions

Investigating the behaviour of the culture raises a whole series of questions. For example—when the culture is disembodied, no sensory signals are being input, yet neurons within the culture still appear to be firing in an occasional structural way. Unfortunately connecting electrodes into the culture in order to directly measure the signals would then affect the culture and, in a sense, embody it. Questions could be asked as to what does its body mean to the culture? Or who or what does the culture think it is? An attempt will be made to deal with such issues in the following section.

It is quite possible, as an alternative, to employ human neurons rather than rat neurons as the decision making brain of the robot. Technically this presents a few different challenges, however, it possibly presents more of an ethical rather than a technical problem. It is hoped that any results obtained in embodying cultured human neurons within a physical robot body will produce much more meaningful results in terms of studying human neural conditions and perhaps gaining an understanding of several mental conditions (Aziz 2009).

As with rat neurons, human neurons can be readily obtained from embryos and cultured after dissociation. The use of human neurons does though throw up a plethora of other possibilities and questions. For a start, rather than obtaining the neurons from embryos, humans could be willing to donate their own neurons—either before or after death. Would not an individual like to live on in some form at least, in a physical robot body? On top of this, human neurons would not necessarily have to be dissociated; they could be fed and laid out on the electrode array as slices. In this case it would be interesting to see if some aspect of memories or behaviour remained, if experiences of the brain in a human body stayed to some extent at least.

Perhaps it would be a way of keeping hold of a loved one who became seriously ill. Indeed, if we are looking forward to a time when humans, as old people, have robots looking after them around the home—wouldn't it be far better for the robot to actually "know" its housemate. So if a loved one is soon to die, perhaps scientists could take away neuron slices, culture them and return them as the brain of a brand new household robot. Maybe the robot would exhibit some of the emotional tendencies and traits from the loved one that would bring back happy memories. But for human neurons, with some sort of awareness of their new existence, how would old memories sit with this? Would it indeed be too traumatic an experience?

Robot consciousness

Clearly we cannot go far in our investigation of culturing robot brains before we need to ask the question as to whether or not the brain experiences consciousness. At present a typical brain of this type, splayed out on a 2-Dimensional array, contains around 100,000 neurons. Nothing like the 100 billion neurons (typical) housed in a human brain. So for those who feel size is important then maybe consciousness cannot yet be considered.

But new lattice culturing methods are being investigated which allow for a 3-Dimensional culture to be kept alive and grown. When we consider a 3 Dimensional volume brain being kept alive and embodied, this means we are looking at a robot brain with (typically) 30 million neurons. In fact, looking ahead a little, a 4,000 × 4,000 2-Dimensional structure of 16 million neurons would result in a 3 Dimensional volume brain consisting of over 60 billion neurons—more than half the size of a typical human brain and—given the typical in situ human neuron death rate—possibly not far away from that of an elderly human. Probably such a robot brain would in fact be more powerful (in terms of numbers of neurons and connectivity) than that of a stroke patient for whom a whole section of their brain has experienced neuronal death.

There are many different philosophical arguments as to the nature and extent of consciousness. For example there are those who feel that it is a unique quality apparent in the human brain (Penrose 1995), whereas others believe it is a property of all biological creatures and that the neurons of other animals have the same functionality as human neurons (Cotterill 1997, 1998). It would not be possible to go into considerable depth in such an article as this. Rather an attempt has been made to use the technology discussed to raise questions as to what each person holds as a concept of consciousness.

So how do we now consider the consciousness of our robot when it has a brain which consists of 60 billion densely packed, highly connected and developed human neurons? Can we endow it with genuine understanding and, as related by Penrose (1995), therefore genuine intelligence. If so, we will definitely have to think about giving the robot voting rights, allowing it to become a politician or a philosophy professor if it wants to and the possibility of putting it in prison if it does something it should not.

But what are the arguments against our robot being conscious? Could it be that 60 billion is still not 100 billion and that's all there is to it? If so, then maybe we will need to start counting the number of brain cells in each human's head such that those whose total falls below a threshold (let's say 80 billion) will find themselves dropped from the human race on the grounds that they are no longer a conscious being. Perhaps we will need some basic test of communication such as the Turing Test (Turing 1950) and everyone must achieve a basic standard in order to avoid the cut and what—be incinerated?

The simple fact is that the best communication machines are now knocking at the door of passing the Turing Test (Lafsky 2009). Surely our robot with 60 billion brain cells would be able to get somewhere close—maybe even performing considerably better than some humans. If so, who would we be to deny the robot its own life?

So could it be emotional responses that are important instead? But if the robot has human neurons couldn't it potentially experience similar (if not the same) emotions to humans? It might be argued that the mere fact that human neurons are involved is not in itself sufficient for our robot to exhibit an identical form of consciousness to that of a human. The point is here though that it is up to an objector to say what extra they believe is required. In any case are we actually interested in an identical form of consciousness to that of a human or rather just some form of consciousness?

Perhaps our robot must have the same sensory input as humans to be deemed conscious? Well even now audio input abilities are being given to the robot, olfactory (smell) is another short term possibility along with basic touch and vision systems. The only difficulty appears to be with taste, perhaps due to its extreme subjectivity. But there again there are many individual humans who have a very poor (or no) sense of taste. Yet surely we would not suggest that all people who have no sense of taste are not conscious—or those who are blind or have a hearing deficiency for that matter. Clearly sensory input in itself is not critical to one's status as a conscious being.

Even more contentious would be an argument suggesting that motor skills are important to consciousness. The present robot moves around quite quickly on wheels. Most humans move around on two legs and manipulate with two arms. But some humans move around on wheels—in fact the world record for the marathon is held by a wheelchair athlete rather than a biped. Meanwhile other humans have no arms or, in a few cases, have robot arms. Then there are those who have contracted Motor Neuron Disease and have limited movement abilities due to a malfunction in that specific part of their brain. It would be horrendous to suggest that any such humans, e.g. Stephen Hawking, are not conscious beings. Clearly motor skills cannot be considered as a tester for consciousness. In any case we are presently embodying a culture in a biped walking robot body, with arms and hands that can grasp and pick up—so overall the robot may well soon have better performance abilities in this area than many humans.

What we are faced with therefore is an entity with a robot body and a brain consisting of human brain cells. Because the creature has a physical robot body is, I contest and have discussed, not a sufficient reason to claim that it is not a conscious being.

Surely, because I call it a 'robot' is also not a creditable reason for anyone to deny that it is a conscious being. Otherwise I could start referring to the entity as a 'human' and we would, on that basis, have to agree that it is conscious. In other words everyone would be deferring their considered judgement on an entity's consciousness to what I might happen to call it. On the other hand, if I started to refer to another human as a robot then, by the same argument, that human would not be conscious because of what I was calling them. So their opinion wouldn't matter anyway! Whilst it is nice momentarily to hold such a powerful position on the life and rights of others, I would suggest that what I/we call an entity is also not a sufficient classifier on which to base our decision on its consciousness.

The value of an education

In our quest to find a reason why our robot, with a brain developed from human neurons, cannot be granted the status of a living, conscious being, with all the opportunities, protection, rights and laws that such status endows, I have explored here and, I believe, dismissed many potential physical (including neural) stumbling blocks. Indeed it is difficult to suggest any further possibilities on this front.

In this discourse I will not waste time to consider a mystical or magical solution to the problem. Included in this is the 'God' syndrome (by this is meant the possible argument that God created humans but did not create our robot—therefore all scientific arguments are null and void). The basis apparently being that—we cannot find any reasonable scientific answer therefore we will put it all down to either God or some magic dust.

The same type of argument can employ some basic difference as being the defining issue, no matter how irrelevant this difference might be (we will look at the functionality argument later). This argument was used merely a century ago to supply racial material in an attempt to prove that black humans were not as intelligent as white humans, e.g. because their sexual organs are a different distance from their naval (Warwick 2001). Such arguments have no known scientific basis and will not be entertained here. In exactly the same way, it is not appropriate to deny that our robot is conscious simply because it looks different. In fact by linking our own robot brain with a presently available animated humanoid robot body, it may well then look and act in the same way as many humans do.

It might be argued that human consciousness may emerge (in absolutely all humans) in a qualitatively different way to that of our robot. But as our robot has human neurons as its brainstock, then given the variety of humans that exist with different mental conditions, who can deny that our robot would not have its own qualitative development and therefore consciousness—or is it simply a case that if we take the robot to the opera a couple of times that will be sufficient? Clearly qualitative development is not relevant and I will waste no further time on the issue.

What we do appear to be left with are the two critical properties of nature and nurture—arguably the basic elements of human intelligence. Are we going to deny our robot its consciousness because of its educational background? It did not have the appropriate experiences or perhaps it didn't go to the right school—therefore it is not a conscious being—maybe on this basis the English class system is still thriving! Again we would have to start looking at the education of humans and deny some the basic rights of others because they went to the 'wrong' school. Such an approach would be, dare I say, somewhat contentious. Quite simply education/nurture cannot be used as a basic argument against our robot's consciousness. In fact, even the present robot, as it is moving around in the lab, is obtaining a University education of sorts.

So what we appear to be left with is the concept of nature. How an entity comes into being must be important as a decision making tool as to whether or not that entity is conscious. It doesn't matter what we call it, it doesn't matter how it senses the world around it or how it interacts with its environment, it doesn't matter what education it received, and so on. All that can be important is how it came to life. If this is not the important issue then surely we will have to admit that the robot is conscious.

But even here we have problems of drawing the line. It has to be said that at present it does not look to be possible to bring such a robot to life through some form of sexual act between two humans. But today we must also allow for, and take in to our discussion, techniques such as test tube babies for humans and even cloning. However, it must be realised here that the human neurons which actually constitute the brain cells of the robot came about in one of these manners—very likely in fact through the relatively straightforward sexual version.

Discounting educational and environmental effects, the only difference between the robot brain and a human brain might therefore be deemed to be merely down to the length of gesticulation—but this would seem to be an extremely weak line to draw for a strong division in decision making with regard to an entity's state of consciousness—especially when we consider the situation of premature babies.

Variety of humans

Perhaps the case for our robot with human neurons has been made, but possibly it is not watertight, maybe there is a loophole or two. What the argument does throw up though are innumerable questions regarding how we consider other (non-robot) humans and in particular extreme cases, such as those on life support mechanisms or those affected by dementia. Because our consideration of human consciousness, with its knock on effect of awareness and rights, must necessarily apply to all humans, it is not merely applicable to philosophy or IT professors.

The point here is that it is extremely difficult, if not impossible, on any scientific basis, to exclude our robot from the class of conscious entities. On top of this, because its brain is made up of only human neurons it is extremely difficult to find grounds on which to discriminate against it, particularly when it may well be, in some ways, nearer the human norm than some disadvantaged human individuals.

Chinese room

There may be some who feel that if the Turing Test can't come up with a solution then may be the Chinese Room can (Searle 1997). But whether the Chinese Room argument holds water or not, the logic employed in the Chinese Room argument is cemented on the basis that human brains are different from computer/machine brains due to the 'emergent property' of the human brain—any conclusions drawn are then focussed on the assumption that human brains appear to have something extra in comparison with machine brains. Our robot though does not have a digital/computer/machine brain, rather, just like you and I, it has a brain full of biological neurons—soon to be human neurons. If we can conclude anything at all from Searle's Chinese Room argument it is that our robot is indeed conscious.

In fact Searle (Searle 1997) stated that "The brain is an organ like any other; it is an organic machine. Consciousness is caused by lower-level neuronal processes in the brain and is itself a feature of the brain." Searle also talks of an emergent property, which implies the more neurons there are, with greater complexity, so this eventually results in the form of consciousness exhibited by humans. As our robot, we are assuming, will, in time, have a brain consisting of several billion highly connected human neurons then by Searle's argument we must assume that it will have a form of consciousness that is pretty much on terms with humans, whatever its physical embodiment.

The point is here not that I am claiming that the emergence of some form of consciousness depends on the size of the brain and the type of the neurons, but that at least one philosopher (Searle 1997) points to that conclusion. For you, the reader, to deny that our robot exhibits some form of consciousness at least, then you need an alternative (scientifically based) argument.

Functionality argument

One argument you might have follows the line taken by Cotterill (1997, 1998), with its extension into embodied robotics by Asaro (2009). In this case it could be argued that what actually matters in terms of consciousness is the functional organization of neural cells and not merely the quantity. The argument could continue that a dish containing 200 billion neurons has no better chance of being conscious than a dish of 20 neurons and that the interesting properties of neurons derive from their associations not their quantity nor their quantity of associations—what matters are the functional structures formed through these associations.

It could be argued that even with a $4{,}000 \times 4{,}000 \times 4{,}000$ culture of neurons it is not clear that this could presently be grown to be functionally equivalent to even part of the functioning human brain, e.g. the hypothalamus. On top of this, such structures in the human brain use a multitude of electro-chemical communication channels that are not realisable with the present day multi electrode arrays. Further, there are many different types of neural cells and support cells and that some combination of these different types of cells might be essential for consciousness.

This is indeed a strong argument, but ….

Firstly, of the two paragraphs forming the argument I have no dispute with the entire contents of the second paragraph. In this case it is certainly true that with present day knowledge it would be difficult to imagine realising anything that was a rough copy of part of the human brain in its functioning. This said, as the robot brain develops, even in the two dimensional case, so neurons do indeed appear to take on different roles—motor, sensory, routing, support etc.—but these roles and their performance are possibly different to those in the human brain.

But it has to be said that we are not trying here to achieve a form of intelligence that is an exact copy of human intelligence, or a form of consciousness that is an exact copy of human consciousness, but rather to consider the possibility of our robot being intelligent and conscious in its own right and way. The fact that our robot brain does not work in exactly the same way as a human brain does is therefore only relevant to the argument if it is 'definitely' the case that such differences are critical to the existence of consciousness in any form.

All I am saying here is that our robot 'could' be conscious in some way, not that it 'definitely' is conscious. If you say that such differences might or might not be relevant (and not that they 'definitely are' relevant) then you must agree with the point that our robot 'could' be conscious. If however, you say that such differences 'definitely' are relevant then this means that you have proven scientific evidence (not that you would simply like it to be the case) and that, as Penrose (1995) put it, you know the "essential ingredient … missing from our present-day scientific picture". I personally am not aware that such knowledge, regarding the existence of consciousness, exists at present.

As far as the first paragraph used in the functional argument is concerned—I would argue that the number of neurons IS important as indeed are the quantity of connections. However, I agree that this size is probably not the be all and end all, and that functional organisation is also critical. However, as for the dish of neurons—well—if we consider a creature with 20 neurons in their brain—a type of sea slug maybe—then by the same argument the sea slug MUST be conscious, otherwise humans (with 100 billion neurons) could not be conscious, as the fact that humans have far more neurons is not, so the argument goes, relevant to consciousness. Personally I feel that humans are most likely conscious in a much more profound way than a sea slug and that by the same token a dish of 200 billion neurons (if it is conscious) will be conscious in a much more profound way than a dish of 20 neurons.

Case for rights

All of this brings us on to some key issues. At present with 100,000 rat neurons, our robot has a pretty boring life doing endless circles around a small corral in a technical laboratory. If one of the researchers leaves the incubator door open or accidentally contaminates the cultured brain then they may be grumbled at and have to mend their ways—but that's all. No one faces any external inquisitors or gets hauled off to court.

With a robot whose brain is based on human neurons, particularly if there are billions of them, the situation might be different. The robot will have more brain cells than a cat, dog or chimpanzee. To keep such animals there are regulations, rules and laws. The animal must be respected and treated reasonably well at least. The needs of the animal must be attended to—they are taken out for walks, given large areas to use as their own or actually exist (in the wild) under no human control. Surely a human neuron robot must have these rights and more? Surely it cannot simply be treated as a 'thing' in the lab? Importantly, if the incubator is left open and the robot dies (as defined by brain death) then someone must be held responsible and must face the consequences.

We need to consider what rights such a robot should have. Do we also need to go as far as endowing it with some form of citizenship? Do we really need to protect it by law or is the whole thing simply a bunch of academics having some fun? Clearly if you are the robot and it is you who have been brought to life in your robot body, by a scientist in a laboratory, and that scientist is in complete control of your existence it must be an absolutely terrifying experience. It may not be very long before such robots are actually brought into being—such a situation will therefore be apparent in the not too distant future. As a scientist would it be acceptable, as it is now, for me to quite simply take the life of a robot when that robot has a brain which consists of 60 billion humans neurons, if I so wish?

Corollary

For some reason the topic of Artificial Intelligence, in its classical mode, was concerned with getting machines to do things that, if a human did them they would be regarded as intelligent acts (Minsky 1975). That is, AI was all about getting machines to copy humans. There are indeed still those who feel that this is indeed what the subject of AI is all about.

Such a limited view presented, to many, well defined bounds which considerably restricted both technical and philosophical development. Unfortunately significant philosophical discussion has subsequently been spent/wasted merely on whether or not silicon brains could ultimately copy/simulate human brains, could they do all the things that human brains do, could they be conscious (as a human)? The much more important topic of considering the implications of building machine brains which are far more powerful than human brains has, by many, been tossed aside as being merely in the realms of science fiction. What a shame! This is a much more interesting question because it points to a potential future in which intelligent, and possibly conscious, beings can outthink humans at every turn. If such entities can exist then this could be extremely dangerous to the future of humankind.

The size of the cultures employed thus far for neuron growth has been restricted by a number of factors, not the least of which is the dimensional size of the multi electrode arrays on which the cultures are grown. One ongoing development at present is enlarging such arrays for more detailed studies, not only providing more input/output electrodes but, at the same time, increasing the overall dimensions. If this increase in size is mapped onto a lattice structure then things move on apace with regard to the size of individual robot brain possible.

A 300 $\times$ 300 neuron layout results in a 90,000 neuron culture when developed in 2 Dimensions (at the smaller end of present day studies) and this becomes 27 million neurons in a 3 Dimensional latticed structure. Meanwhile a 400 $\times$ 400 layout achieves a 160,000 culture in 2 Dimensions (at the top end of present day studies) and this becomes 64 million neurons in a 3 Dimensional latticed structure. But if this is pushed forward to a 5,000 $\times$ 5,000 neuron layout, it results in a 25 million culture in 2 Dimensions (which undoubtedly we will witness before too long), which becomes 125 billion in a 3 Dimensional lattice. It is not clear why things should stop there however. For example, moving forward to a 7,500 $\times$ 7,500 layout, in 2 Dimensions this achieves a 56.25 million culture which becomes 421 billion neurons in 3 Dimensions—an individual brain which contains four times the number of (human) neurons as are contained in a typical human brain.

Drawing conclusions on developing robot brains of this size, based on human neurons is then difficult. There are certainly medical reasons for carrying out such research—for example to investigate the possible effects on such as Alzheimer's disease by increasing the overall number of useable neurons. But this approach neglects to consider the repercussions of bringing into being a brain which has the potential (certainly in terms of numbers of neurons) to be more powerful than the human brain as we know it.

The purpose of this paper has been to consider the role of biological brains within the field of artificial intelligence and to look at their impact on some of the discussions, particularly with regard to consciousness, that have taken place. Many books have been written on these subjects and hence it is clearly not possible to cover anything like all aspects in one paper. It has not been the case that I would wish to claim that such a brain is definitely conscious, in some sense, but rather to consider how different concepts of what consciousness is, deal with this type of brain. Each person has their own views on what consciousness is and what it is not—I therefore leave it up to you to consider how your own viewpoint is affected (if at all) by the consideration of such brains. Is our robot brain conscious?

Acknowledgments I would like to express my enormous gratitude to the team at University of Reading, on whose considerable work this article is based. Ben Whalley, Slawek Nasuto, Victor Becerra, Dimi Xydas, Mark Hammond, Julia Downes, Matt Spencer and Simon Marshall all deserve rich rewards. The practical work described in this article is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant No. EP/D080134/1. I also wish to mention the unnamed reviewers for this paper who did a thorough job. This led to several improvements to the paper including the addition of the section entitled "Functionality Argument".

References

Arkin, R. (2009). *Governing lethal behaviour in autonomous robots*. London: Chapman and Hall.

Asaro, P. (2009). Information and regulation in robots, perception and consciousness: Ashby's Embodied minds. *International Journal of General Systems, 38*(2), 111–128.

Aziz, T. (2009). Personal communication.

Bakkum, D., Shkolnik, A., Ben-Ary, G., Gamblen, P., DeMarse, T., & Potter, S. (2003). Removing some 'A' from AI: Embodied cultured networks. In *Proceedings of the Dagstuhl conference on embodied artificial intelligence* (pp. 130–145).

Brueckner, A. (1986). Brains in a vat. *Journal of Philosophy, 83*(3), 148–167.

Cotterill, R. (1997). On the mechanism of consciousness. *Journal of Consciousness Studies, 4*(3), 231–247.

Cotterill, R. (1998). Enchanted looms: Conscious networks in brains and computers. Cambridge: Cambridge University Press.

DeMarse, T. B., & Dockendorf, K. P. (2005). Adaptive flight control with living neuronal networks on microelectrode arrays. In *Proceedings of 2005 IEEE international joint conference on neural networks* (pp. 1549–1551), Montreal.

Hebb, D. (1949). *The organisation of behaviour*. New York: Wiley.

Kafka, F. (1972). The metamorphosis. Bantam Classics.

Lafsky, M. (2009). How can you tell if your IM buddy is really a machine? Discover Magazine, 23rd March 2009.

Lewicki, M. (1998). A review of methods for spike sorting: The detection and classification of neural action potentials. *Network: Computation in Neural Systems. 9*(4), R53–R78.

Lloyd, D. (1991), Leaping to conclusions: Connectionism and the computational mind. In T. Horgan & J. Tienson (Eds.), *Connectionism and the philosophy of mind* (pp. 444–459). London: Kluwer.

Marks, P. (2008), Rat-brained robots take their first steps. *New Scientist, 199*(2669), 22–23.

Minsky, M. (1975). *The psychology of computer vision*. New York: McGraw-Hill.

Moravec, H. (1990). Mind children: The future of robot and human intelligence. Harvard: Harvard University Press.

Penrose, R. (1995), Shadows of the mind. Oxford: Oxford University Press.

Potter, S., Lukina, N., Longmuir, K., & Wu, Y. (2001). Multi-site two-photon imaging of neurons on multi-electrode arrays. *In SPIE Proceedings, 4262*, 104–110.

Searle, J. (1997), The mystery of consciousness. New York Review Book.

Shelley, M. W. (1831). Frankenstein or the modern prometheus. London: Colburn & Bentley.
Shkolnik, A. C. (2003). Neurally controlled simulated robot: Applying cultured neurons to handle an approach/avoidance task in real time, and a framework for studying learning in vitro. Masters Thesis, Department of Computer Science, Emory University, Georgia.
Thomas, C., Springer, P., Loeb, G., Berwald-Netter, Y., & Okun, L. (1972). A miniature microelectrode array to monitor the bioelectric activity of cultured cells. *Experimental Cell Research, 74*, 61–66.
Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.
Warwick, K. (2001), QI: The quest for intelligence. Piatkus.
Warwick, K. (2003), Cyborg morals, cyborg values, cyborg ethics. *Ethics and Information Technology, 5*, 131–137.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Andrews, B., et al. (2003). The application of implant technology for cybernetic systems. *Archives of Neurology, 60*(10), 1369–1373.
Warwick, K., Gasson, M., Hutt, B., Goodhew, I., Kyberd, P., Schulzrinne, H., et al. (2004). Thought communication and control: A first step using radiotelegraphy. *IEE Proceedings on Communications, 151*(3), 185–189.
Warwick, K., Xydas, D., Nasuto, S., Becerra, V., Hammond, M., Downes, J., et al. (2010). Controlling a mobile robot with a biological brain. *Defence Science Journal*, *60*(1), 5–14.
White, R., Albin, M., & Verdura, J. (1963). Isolation of the monkey brain: in vitro preparation and maintenance. *Science, 141*(3585), 1060–1061.
Xydas, D., Warwick, K., Whalley, B., Nasuto, S., Becerra, V., Hammond, M., et al. (2008). Architecture for living neuronal cell control of a mobile robot. In *Proceedings of European robotics symposium EUROS08, 2008* (pp. 23–31). Prague.