Black boxes on wheels: research challenges and ethical problems in MEA-based robotics

Bentzen, Martin Mose

Published in:
Ethics and Information Technology

Link to article, DOI:
10.1007/s10676-016-9415-z

Publication date:
2016

Document Version
Peer reviewed version

Citation (APA):
Bentzen, M. M. (2016). Black boxes on wheels: research challenges and ethical problems in MEA-based robotics. *Ethics and Information Technology*, *19*(1), 19-28. https://doi.org/10.1007/s10676-016-9415-z

---

Ethics and Information Technology manuscript No.
(will be inserted by the editor)

Black Boxes on Wheels

Research Challenges and Ethical Problems in MEA-based Robotics

Received: date / Accepted: date

Abstract Robotic systems consisting of a neuron culture grown on a multielectrode array (MEA) which is connected to a virtual or mechanical robot have been studied for approximately 15 years. It is hoped that these MEA-based robots will be able to address the problem that robots based on conventional computer technology are not very good at adapting to surprising or unusual situations, at least not when compared to biological organisms. It is also hoped that insights gained from MEA-based robotics can have applications within human enhancement and medicine. In this paper, I argue that researchers within this field risk overstating their results by not paying enough attention to fundamental challenges within the field. In particular, I investigate three problems: the coding problem, the embodiment problem and the training problem. I argue that none of these problems have been solved and that they are not likely to be solved within the field. After that, I discuss whether MEA-based robotics should be considered pop science. Finally, I investigate the ethical aspects of this research.

Keywords MEA-based robotics · neuron culture · coding · embodiment · training · research ethics

1 Introduction

The possibility of using biological neurons grown on microelectrode arrays (MEAs) as part of robotic systems has been studied for approximately 15 years, see DeMarse et al (2001), Cozzi et al (2005), Novellino et al (2007), Warwick et al (2010b), Tessadori et al (2012). It is hoped that MEA-based robotics will address the well-known problem that robots based on conventional computer technology are not very good at adapting to surprising or unusual situations, see Tessadori et al (2012). One could compare the clumsy and irregular movements of, e.g., most current vacuum cleaner robots to the rapid escape response through jet propulsion of the squid or the sophisticated wall-following behavior of the cockroach, see Dickinson et al (2000), Cowan et al (2014). It is also hoped that biological AI will have applications within human enhancement and medicine, both directly by enabling us to design neural prosthetics and indirectly by teaching us more about the biological principles of the brain. Researchers within the field are not modest about what they have achieved. In Warwick et al (2010b), the researchers claim that ‘research is ongoing in which biological neurons are being cultured and trained to act as the brain of an interactive real world robot’. The title of the paper is ‘Controlling a mobile robot with a biological brain’ and one of the keywords is ‘intelligent controlling mechanism’. Further, the researchers claim to be able to ‘train’ the neuron culture. In Tessadori et al (2012), it is claimed that ‘these results prove that an in vitro network of biological neurons can control an external agent’. There are several questionable terms employed above, amongst which are ‘brain’, control’ and ‘intelligent controlling mechanism’, as well as ‘training’ a term that implies teaching goal-directed behavior. To take an example, is it correct to call a neuron culture grown on a two-dimensional array a brain? The brain of an actual evolved biological organism is a highly complex, functionally organized, three-dimensional object, far from the much simpler two-dimensional cultures grown for the purpose of current MEA-based robotics - although even these simpler cultures are very hard to understand.

In this paper, I first provide a brief history of MEA-based robotics to situate the field in a scientific context. After this, I present and analyze three principal research challenges for MEA-based robotics related to neural coding, embodiment, and the training of neuron cultures. I consider whether MEA-based robotics should be considered pop science or proper science. Finally, I take a look at the ethical aspects of

---

Address(es) of author(s) should be given

this research. Some researchers within the field have claimed that their research provides an ethical advantage, see Warwick et al (2010b), Warwick (2010), Warwick (2012). Supposedly, this perceived advantage comes from a comparison to research that uses actual animals to control robotic body parts, as reported in, e.g., Reger et al (2000) and Talwar et al (2002), as the latter research is more invasive. I will argue that, even if we find this comparison acceptable, the research introduces several hitherto unknown ethical complications into current robotics, which makes it hard to speak of an advantage.

2 A brief history of MEA-based robotics

It has been known since the beginning of neuroscience that neurons communicate through both electrical and chemical signals, see Pereda (2014). Hence, the wish for establishing direct electrical contact with neuron cultures through stimulation and recording of signals arose rather naturally. To avoid various problems with the placement of electrodes in live animals, the idea of growing neurons directly on a microelectrode array (MEA) was conceived. Neurons are taken from live animals (often embryonic rats) and placed in a thin layer on an MEA. After a few days the neurons start to network and exchange chemical and electrical signals. Via the electrodes on the array, researchers can stimulate selected areas of the neuron culture with electrical signals as well as record the electrical output coming from it. Fairly reliable input-output relations can be established, called *neural pathways*. Apart from the signals induced by the input, spontaneous local electrical activity can be recorded at various output electrodes, as well as global activity of a short duration spanning the entire network, referred to as *bursts*, see Maeda et al (1995). The first reported development and study of microelectrode arrays is Thomas et al (1972). However, this pioneering study failed to make recordings, a goal that was accomplished in the study reported in Pine (1980). The first study to use MEAs as a controlling mechanism for a virtual robot (more precisely a virtual animal or animat) is DeMarse et al (2001). It was hoped that the new research object could help us understand cognition, create novel neural prosthetics and foster new forms of artificial intelligence, see Bakkum et al (2004). The researchers succeeded in creating a system that could obtain information from a living neural network, using this information to control the animat through mediating software and stimulating the network. The researchers failed to determine how or if the visual patterns produced by the animat were connected to the neural coding that is supposedly present in the neural network. We call this problem *the coding problem*. At the core of the coding problem is communication. Do we know the language of the neuron culture, can we understand signals coming from it and can it in turn understand our signals? The animat also did not display goal-directed behavior; in other words, it was unlikely that the neuron culture had any intentional control over its virtual robot body. We call this *the embodiment problem*. The embodiment problem is closely related to control. What does it take for us to say that the neuron culture is, functions like or behaves as if it were a brain controlling a body? Finally, the researchers did not determine how to teach the neuron culture how to improve its performance. We might call this the *training problem*. The core of the training problem is desired change. How can we get the neuron culture to develop in a direction we would like it to? The three problems are interconnected, and researchers usually claim to address several or even all three problems at once, see Martinoia et al (2004), Cozzi et al (2005), Novellino et al (2007), Warwick et al (2010b), Tessadori et al (2012). However, for the purpose of understanding the limitations of the field, it is analytically more appropriate to separate the three problems, which I will do in the following section.

3 Research challenges

3.1 The coding problem and communication

Although it is widely acknowledged that neuron communication is strongly related to electrical activity, the message from experts in neuroscience is quite clear: at the moment, we know very little about decoding or reading the neural code (i.e., how to translate signals from neurons into information about perception or behavior) and even less about encoding or writing the neural code (i.e., how to translate sensory information into signals understood by neuron cultures), see Stanley (2013).

> Presently, we do not understand the nature of the coding systems that are used in single neurons, and it is unclear what sorts of dimensional reduction are possible across populations and networks of neurons.

Cowan et al (2014)

Because this is clearly the case, how do the researchers in MEA-based robotics obtain results? Would it not be impossible to induce perception about the external environment in the neuron culture and to interpret the signals from the neuron culture as motor commands without a basic understanding of this neural code? Although the answer seems to be a clear ‘yes’, it is instructive to observe how they proceed to obtain results. First, some background: According to the neuroscientific paradigm, all information about the sensory world comes from an organism’s observation of its own neurons in real time, see Bialek et al (1991). Neurons send out electrical pulses referred to as action potentials or spikes, which are regarded as essentially identical. Sequences of spikes are referred to as spike trains. It is believed that the neural code is intrinsically linked to these spike trains. However, there is a long standing dispute about whether the neural code is based upon the mean firing rate of spikes (a rate code) or whether the spacing between spikes also carry information (a timing code), although it has also been suggested that both coding schemes might be active at different time scales, see Stanley (2013). In the early 1990s, Bialek and collaborators developed an important model of information transmission from the sensory system to the central nervous system. The model was based upon experimental studies of a movement-sensitive neuron in the visual system of the blowfly *Calliphora erythrocephala*, and similar studies, see Bialek et al (1991). The researchers were able to reconstruct the waveform of the time-variant visual stimuli through an algorithm that takes the spike trains of a neuron as input. In other words, they could reconstruct some of what the fly sees through its neural activity. The approach is justified through matching of the resultant waveform with the known waveform of the stimuli. A simplified model of how to reconstruct stimuli from spike trains was suggested in Gabbiani and Koch (1996). This model was in turn used in Novellino et al (2007) in reverse order in an application within MEA-based robotics. Sensory input from a robot was averaged and presented to the neuron culture as spike trains. The output from the culture was translated via conventional algorithms and presented to the robot body as motor commands. However, there is a fundamental problem with this approach. The work of Bialek and successors show using methods from information theory and probability theory that it is possible to encode information via spike-trains. However, this result leaves the question as to whether the organism actually does encode information in this way completely open.

> . . . the reconstructions performed on H1 [the neuron] in the house fly and on cells in other animals, as well as this theoretical work, leave totally open the important problem of determining whether the information on a time-varying stimulus that can be encoded in a neuronal spike train is actually used by the organism . . .

Gabbiani and Koch (1996)

The researchers in MEA-based robotics base their research on assuming that a previous result holds beyond its boundary. They assume knowledge about writing the neural code. In fact, they are encoding information in a way that might not be decoded in any way. First, because the neuron code might be fundamentally different from what is assumed. For instance, current research suggests that the communication between neurons is more complicated and depends heavily on synchrony between neurons on a timescale of a few milliseconds, Stanley (2013). Taking this into account would require a refinement of the present recording and stimulation protocols. Further, because of the embodiment problem discussed below, researchers within this field do not have any reliable biophysical information to check whether the assumptions about coding are justifiable. Because there is most likely no coherent organism whose behavioral data can be measured, the research cannot contribute to solving the coding problem.

3.2 The embodiment problem and control

The basic idea behind MEA-based robotics is to provide a neuron culture with a robot body. The goal of this research is to create a hybrid being in control of its own behavior. The challenges involved in reaching this goal we call the embodiment problem. It should be mentioned that it is in principle possible to solve the embodiment problem independently of the coding problem. That is, the neuron culture could in principle be in control of a robot body without us knowing the exact coding mechanism enabling it to control its body parts. However, because the actions of current MEA-based robots rely heavily on researchers interpreting neural activity from the network and translating it into motor signals to the robot body, it is not very likely that the embodiment problem will be solved before the coding problem. In fact, it is recognized that

> ...one still knows very little about the fundamental neuronal processes that give rise to meaningful behaviours...

Warwick et al (2010a)

Because this is the case, we can ask a similar question to the one we asked in the previous section. How do MEA-based robotics obtain results if they really have very little idea of how a neuron culture controls a body? As one study admits about translating neuron activity into motor commands under such conditions

> Any ‘decoding’ strategy is clearly arbitrary...

Martinoia et al (2004)

The answer is that they use the information they obtain for the neuron culture in a way that fit the model of behavior they have chosen. In a number of applications, the goal of researchers has been to make the robot behave as a simple evasive Braitenberg vehicle, see Braitenberg (1986). Instead of waiting for behavior or control to emerge, this means that

> ...outputs were chosen [...] in order to result in a behavior that is as close as possible to the obstacle avoidance. Obstacle avoidance can be achieved if the activation of sensors on one side elicits a decrease of speed at the opposite side.

Cozzi et al (2005)

All it takes to establish a simple negative correlation of this kind is finding an input-output relation where an increase in stimulus causes an increase in recorded neural activity, see Cozzi et al (2005). The neuron culture simply acts as a mere channel through which sensory information flows and is turned into behavior by conventional software. As one study bluntly puts it:

> In the initially developed culture, we found, by experimentation, a reasonably repeatable pathway in the culture from stimulation to response. We then employed this to control the robot body as we saw fit.

Warwick et al (2010a)

Researchers within this field apparently do not fully recognize how problematic this approach is. In MEA-based robotics, the behavioral data, i.e., the movements of the robot, are unreliable as a way of finding out about the principles governing the neuron culture. In spite of a superficial resemblance, the data collected from the movement of the robot body is not comparable to the data collected from a freely moving animal whose brain activity we simultaneously measure. The robot behavior is in fact constructed by conventional algorithms and software to a degree, where we could remove the neuron culture from the robotic system without making much of a difference (except making it more efficient). This is clearly not the case when experimenting with live animals, where the removal of the brain would also stop the animal from moving. The decision mechanism of the robot is manufactured by the researchers and set up on a pragmatic basis as they 'see fit'. This makes it hard to see how proper scientific hypotheses about the behavior of the robot and the activity of the neuron network could be formed, which would actually enhance our knowledge about the biological principles of the brain.

MEA-based robotics is a science in its infancy. I have argued that it is misleading to claim that the neuron culture controls the robot body. Scientists within the MEA community claim that they have created a robot body controlled by a neuron culture. To ensure that this is not a mere quibble about words, it would be prudent to clarify the crucial concept 'control' to make a more thorough evaluation of these claims. However, attempting to do so will easily lead us into deep metaphysical discussions about the nature of free will, autonomy, and so on. To sidestep discussions that are not pertinent to the issues in this paper, the optimal strategy seems to be to opt for the weakest possible definition of control that is mostly compatible with MEA-based robotics. If a definition can be found that makes the claim of the scientists true, this claim can be justified, at least as seen from a perspective that is sympathetic to MEA-based robotics. However, if researchers fail to meet the requirements of such a definition, it is safe to say that they have not accomplished what they claim. However, because finding such a minimal definition is an optimization problem that would require a lot of work and which will be complicated by the vagueness of the term 'the project of MEA-based robotics', the approach to be followed here will be different. I will take three conceptions of control off the shelf, which seem relatively sympathetic to the mechanistic view of the mind shared by many researchers in neuro science, one from philosophy, one from the field of neuromechanics, and one from control theory. These conceptions are of decreasing strength in that they require less and less of the controller. The discussion will highlight what is missing from current MEA-based robotics. Finally, I will discuss the difference between control and mere causation, as an identification of the two trivializes the concept of control.

3.2.1 Dennett's definition of control

The philosophical definition to be considered below is the one presented by Daniel Dennett, see Dennett (1984). Although Dennett claims to capture our everyday concept of control with his definition, he also claims that the definition is compatible with more technical definitions found in automata theory and cybernetics, and in general, he takes care to keep his philosophical views compatible with a determinist view of science, which also seems predominant within robotics, neuroscience, and AI, or at least acceptable within these fields as a philosophical position.

Definition 1 (Control - Dennett) A *controls* B if and only if the relation between A and B is such that A can *drive* B into whichever of B's normal range of states A *wants* B to be in. (If B is capable of being in some state *s* and A wants B to be in s, but has no way of putting B in s or making B go into s, then A's desire is frustrated and to that extent A does not control B).

Applied to the present case, the neuron culture controls the robot body if and only if the neuron culture can drive it into a state of turning, speeding up, etc., when the neuron culture wants it to do so. However, we have absolutely no idea what the robot culture wants. It might be a good idea to highlight the extent of our lack of knowledge by a comparison to a study of the wants and needs of biological rats. A recent study shows that biological rats can feel regret, which the researchers contrasts to disappointment and defines as the realization that a worse than expected outcome is due to one's own mistaken action, see Steiner and Redish (2014). This study is conducted in a socio-economic framework (a restaurant row task) and is based on a correlation between the observation of 'regretful' behavior (e.g., rats looking back towards a better but presently unobtainable option) and brain activity in the orbitofrontal cortex of biological rats and the analogy to the human case, where this area of the brain has been documented to be connected to expressing regret. This study requires at least three elements for its success.

1. A functionally organized brain analogous to the human brain.
2. Bodily expressions of what can be perceived as intentional behavior.
3. A clear idea of what is valued by the animal (food), enabling the connection to an economic theory of rational choice.

None of these elements are available to the researchers in current MEA-based robotics. First, the neuron culture is not known to be organized functionally like a mammal brain. Therefore, it is hard to interpret activity in the network in analogy to activity in areas of the human brain. Second, the bodily expressions of the robot are based on measuring outputs from the culture, which are in turn based on neural pathways found prior to the experiments. These outputs are used as 'researchers see fit', i.e., the neuron culture cannot be said to choose its own action in any way. This means that these bodily expressions cannot be used as an independent control of intentional behavior in analogy to human beings (there is no analogy to looking back for an MEA-based robot). Third, with regard to value, although, the neuron culture needs nourishment, there is no indication that it experiences anything like hunger, e.g., studies do not report trying to link the provision of nourishment with a training protocol for neuron cultures. We must conclude that it is very unlikely that the neuron culture controls the robot body in the sense of Dennett's definition.

3.2.2 Neuromechanics

I now turn to a different conception of control, which will be weaker in the sense that it does not require us to be able to talk directly about the mental states (wants) of the controller because it only focusses on purely physical aspects. In the cross-disciplinary field of neuromechanics, the exact way the brain interacts with other parts of the body is studied, see Nishikawa et al (2007). As one of the applications of neuromechanics is the design and control of mobile robots, MEA-based robotics could look to neuromechanics for a foundation. Applying this conception of control to the case of MEA-based robotics will further highlight what is missing from the current research in this area. The following quote shows how motor control is conceived within neuromechanics.

> Motor control fundamentally involves a series of transformations of information among different levels and components of the neuromuscular and skeletal systems. Sensory information (proprioceptive and exteroceptive) is transduced by sensory structures that in turn transfer a subset of their information to the central nervous system which, following yet another transformation, issues a set of motor commands. The motor commands trigger force development in muscles, which drive movement and control the mechanics of the body.

Nishikawa et al (2007)

According to neuromechanics controlled (intentional) movement involves a drive from a higher brain center that is communicated to a lower network. This simple command is transformed by the network into muscle-tendon actions distributed over various limbs and joints. Clearly, important components of this model are missing from the current research setup in MEA-based robotics. The distinction between a higher brain center driving the network and the network is completely missing, making any talk about a *controller* or a *controlling brain* dubious. This is because a neuron culture grown on a two-dimensional MEA is not functionally organized. Only one type of cell is used, neglecting the fact that various cell types are most likely required for functional organization, see Zeisel et al (2015). Further, the information distributed by the nervous system of a live animal in various ways to facilitate various forms of movement is replaced by a simple output, making any talk about *network controlled movement* dubious. This is because the research setup only allows recording of electric output from the neuron culture which is then translated directly into motor commands, neglecting the intermediate role played by the peripheral nervous system in establishing connections between the limbs and the central nervous system. Because of these missing components, I must conclude that according to the current conception of motor control within neuromechanics, there is no such thing as a brain controlling a body in MEA-based robotics. We might also note a further deficit in the research setup, which is that, unlike an engineered device, such as a remote controlled vehicle, an evolved device, such as a hand or a mouth, is typically highly multi-functional and flexible - the human mouth can, for instance, be used to swallow, yawn, bite, talk, and kiss, which are activities that have very different purposes, or muscles, which are complex and versatile devices that contribute a lot to the stability and flexibility of the movement of biological systems, quite apart from neurological input. Much of the flexibility of biological systems presumably comes from this physiological flexibility, which is completely lost by simply combining a mechanical engineered device with a neuron culture, thus undermining the goal of MEA-based robotics as described in the introduction.

3.2.3 Control Theory

It appears that there is an even weaker form of control than the ones considered above, as when we say that the thermostat controls the room temperature, which is not intentional and does not require the drive of a higher brain center. This kind of control is studied in the engineering field of control theory, which addresses the change and stability of dynamical systems, biological or otherwise, through regulation via feedback. In this section, I investigate a control theory perspective of neuromechanics, see Cowan et al (2014), Roth et al (2014), and compare the results to that for MEA-based robotics. According to control theory, regulatory feedback can change the characteristics of a closed-loop system dramatically, making an otherwise fragile system robust, a slow system fast, and so on. Perhaps the neuron culture does control the robot body according to this weak conception of control? Granted that the biological robot does constitute a closed loop and that sensory feedback does regulate the behavior of the robot, the question is what the function of the neuron culture is. The answer seems to be that, in regard to controlling the robot body, the neuron culture is an unnecessary appendix whose removal would only enhance the performance of the system. The information delivered to and from the neuron culture is very coarse grained. On the other hand,

> A hallmark of a high-performance control system is the ability to achieve large responses over a wide range of frequencies in response to stimuli (change) without skirting too close to the instabilities that can result from high-gain, large-latency feedback (stability).

Cowan et al (2014)

It is clear that current MEA-based robots are not high-performance systems and that we are far from achieving the flexibility of biological systems. However, one may argue that the poor performance of these robots does not indicate that there is no control, only that there is poor control because when we are attempting to use a tool that we are unaccustomed to; this argument seems rather to presuppose control than to establish it. The fact is that the source of the stability of the biological robot is not related to the neuron culture but to conventional software and the mechanical parts (including sensors). The issue of how to evaluate change within the culture will be addressed in the following section. If there is any important processing of information going on in the network, it is not transformed into anything we can meaningfully call control of the robot body. I therefore conclude that the neuron culture does not control the movements of robot body according to the conception of control theory, although the entire system consisting of conventional software, mechanical parts and as a small and rather irrelevant part, the neuron culture, does control the body.

3.2.4 Weaker forms of control

I have looked at three conceptions of control of decreasing strength. I cannot completely rule out that an even weaker conception can be found which validates the researchers' claim. When looking for such a definition, however, we should also consider that a definition of control can become so weak that we cannot distinguish control from mere causation. According to most views on causation, e.g., the counterfactual account or the regularity account, the output from the neuron culture does cause the robot body to move, i.e., it is a link in the causal chain leading to the robot body turning, see e.g., Pearl (2009). However, because this output in turn is caused by the input from the robot body, are we then to say that the robot body controls the neuron culture in these cases? If causation is transitive, as at least some philosophers allow, are we then to say that the robot body controls the robot body, that the robot body is in fact self-controlled? Clearly, the identification (or near identification) of causation and control will trivialize the claim that the neuron culture controls the robot body. Finally, we might remark that this section has only been about whether the culture controls the robot body, not whether the culture itself can be controlled, regarding it as a black box, e.g., using methods from fuzzy control theory, see Chen and Ying (1997). It is clear that the entire robotic system is controlled by the researchers, mainly because of the conventional algorithms and mechanical parts such as sensors and wheels. What I have disputed is the claim that the neuron culture controls the robot body. However, when discussing the training problem below, I will consider the possibility of regarding the neuron culture as a black box whose performance we wish to increase.

3.3 The training problem, desired change and learning

There has been a strong focus within MEA-based research on teaching neuron cultures behavior, i.e., training them. At the core of the training problem is the wish to change the neuron culture in ways that researchers desire. This desired change, at least as seen from a network internal point of view, we would call 'learning'. I will first consider a strong concept of learning and then consider a weaker concept in each case evaluating whether it is justified to speak of learning in the context of MEA-based robotics.

3.3.1 Reinforcement learning

A very natural concept of learning is reinforcement learning, rewarding desired behavior and punishing undesired behavior to induce a desired change in behavior. Reinforcement learning is linked to preference, we must know what kind of things the learner prefers to establish a reinforcement learning protocol. However, researchers simply do not know what a neuron culture grown in vitro prefers.

> One major problem with this is deciding what the culture regards as a reward and what as a punishment.

Warwick (2012)
Without knowing what the culture likes or dislikes, if anything, reinforcement learning is ruled out as an option.

3.3.2 Hebbian learning, performance increase and functional plasticity

Reinforcement learning provides a rather strong concept of learning, as it requires making assumptions about the internal states of the neuron culture in terms of its likes and dislikes. As an alternative to reinforcement learning, we might look at the neuron culture as a black box whose output we are attempting to change in a direction we desire. With regard to artificial neural networks, this kind of learning is sometimes referred to as Hebbian Learning, and this term has also been used within the MEA community, see Ferrández et al (2013). In a similar vein, the MEA community has coined the term 'functional plasticity' to indicate those changes in stimulus-response relationships or in spontaneous patterns that are experimentally induced by electrical stimulation and lasting at least on the order of an hour, see Massobrio et al (2015), Wagenaar et al (2006). A standard method of attempting to induce functional plasticity is through what is known as tetanic shocks. This implies giving the neuron culture a 20 Hz stimulation with the subsequent effect of increasing the number of spikes at the output electrode, see Tessadori et al (2012). Taken as a definition of learning, it is clear that functional plasticity is a very weak form, meaning no more than lasting change in the input-output relation. Although we might use such a change to create input-output relations to our benefit, we really cannot say anything about whether that means there is a 'subjective' or 'meaningful' aspect of learning taking place in the neuron culture before we have solved the coding problem or the embodiment problem. As the researchers acknowledge,

> The exact biological mechanisms linking performance increase and tetanic stimulation are still unclear...

Tessadori et al (2012)

If we knew that the desired change was correlated with messages written in the neural code or with meaningful bodily behavior we could get somewhere. In other areas of neuroscience, we have access to reliable behavioral data, giving clues as to how to interpret the neurophysiological activity. Further, the knowledge about the functional organization of a biological brain can also give clues as to how to interpret the interaction between brain activity and behavior. In some cases, it is even possible to open a closed loop to investigate the roles played by specific neuromechanisms, see Roth et al (2014). We have no corresponding brain structure or behavioral data with regard to MEA-based robots. Moreover, it is even controversial whether this weak form of learning, i.e., functional plasticity or performance increase, can even be reliably induced in neuron cultures using current methods, see Van Staveren et al (2005), Wagenaar et al (2006), Massobrio et al (2015), for several reasons. For one, the neuron cultures are spontaneously changeable, which makes it difficult to design a learning protocol with which we can reliably argue that the change is induced and not just happening, see Massobrio et al (2015). Related to this is the fact that the changed (whether induced or not) seem to be of short duration, often disappearing or taking another form between training sessions, see Wagenaar et al (2006). A further methodological problem is how to document changes in the network from the electrical output, Massobrio et al (2015). In short, it is hard to know what kind of output should be interpreted as what kind of change, especially when we do not know what the output means (the coding problem) or even whether the output is actually commands for meaningful actions (the embodiment problem).

In conclusion, the value of being able to teach an MEA-based robot to act in certain ways is indisputable. However, even with a definition of learning that is so weak that it means nothing more than 'lasting change', it is dubious at this point whether cultured networks learn.

4 Is MEA-based robotics pop science?

Given the criticism leveled at MEA-based robotics above, one might be inclined to the conclusion that this is not proper science, but what is derogatorily referred to as 'pop science', a dubious activity aimed more at attracting media attention and funding than the search for truth and respect from the scientific community. However, although this research has received attention from the media, especially through Kevin Warwick's work in the area, this research does not appear to be pop science, at least not according to my investigation of the field. This prima facie conclusion is supported by the fact that the groups conducting this research operate from respectable universities, and many of the pioneering papers have been published in respectable journals. Table 1 shows some pioneering peer-reviewed papers within the field with the impact factor of the journal and the number of citations of the paper. In each case, the journal is relevant for the topic of the paper and is included in the Web of Science. All impact factors, except the one for Ethics and Information Theory, are taken from the 2014 Journal Citation Report Science Edition published by Thomson Reuters as part of the ISI Web of Knowledge. The impact factor of Ethics and Information Technology, Warwick (2010), is taken from the 2014 Citation Report Social Science Edition. The numbers of citations of the individual papers are taken from google scholar and have not been adjusted for self-citations.

Peer reviews, Web of Science, Impact Factors and number of citations are blunt and imprecise instruments for evaluating scientific quality, but they do provide a lower bench-

paper	Journal IF	Paper cited by
Reger et al (2000)	1.386	163
DeMarse et al (2001)	2.066	232
Cozzi et al (2005)	2.083	17
Novellino et al (2007)	0.481	63
Warwick (2010)	1.021	54
Downes et al (2012)	4.620	37
Tessadori et al (2012)	3.568	14

Table 1 Important papers in MEA-based robotics

mark that, in most cases, should save research from deserving the predicate 'pop science.' Nonetheless, the above analysis shows that scientists within this field have been guilty of hasty generalizations and of overinterpreting previous results, treating conjectures as firm ground to build upon, which does not constitute sound methodology. For this reason, the field can appear to have developed further than is in fact the case. To foster real scientific development, future research in this emergent field ought to take a much more critical approach to foundations and to stress limitations and doubts further. Additionally, the relation between everyday words, such as coding, control and learning, and the abstract and simplified versions of these employed in scientific research should be made clear when communicating results more broadly. The latter is a research ethical concern because if the public is misled, it will be to the detriment of science in the long run despite the possibility of fast funding and press coverage in the short run. This leads me to a further discussion of the ethical issues facing MEA-based robotics.

5 Ethical problems

Kevin Warwick has considered ethical aspects of MEA-based robotics in detail in Warwick (2010), Warwick (2012). Warwick acknowledges that robots with biological brains present ethical problems that need to be addressed. Here, I will mainly focus on two claims made by Warwick: First, growing small neuron cultures presents an ethical advantage to experimenting with live animals, and second, growing large neuron cultures of human neurons to study diseases of the brain is ethically acceptable. Warwick and collaborators seem to think that ethical questions related to research using live animals can be solved by using neuron cultures instead, see Warwick et al (2010b). At first, this appears to be an odd comparison to make. Experiments made with live animals have a variety of purposes, some of which are ethically acceptable and some of which may not be. Clearly, experiments with neuron cultures could not replace very many of these, e.g., preclinical tests of drugs on rodents could not be done with neuron cultures instead, as we have to test the effects of the drug on the entire developing animal not just some of its brain cells. Further, even if we only consider experiments made with live animals controlling robot body parts, such as Reger et al (2000), there are fundamental methodological advantages to these experiments, which may well justify them ethically because of the knowledge gained. With live animals, we have a functionally organized brain, and because we already have motor control, it is much more likely that we will be able to replace the controlled body parts or even the entire body with, e.g., robot limbs and form hypotheses about the relationship between behavior and neural activity. There might also be a range of biological questions that can be answered by experiments of this kind, see Roth et al (2014). Even if we grant that this technology presents an ethical advantage because no live animals are harmed during the research process (not counting the pregnant rats and their fetuses whose neurons are used), there are new research ethical questions that will result from this research. Researchers profess that the neuron cultures may have thoughts that are unknown to us - they link these possible thoughts with what we perceive as spontaneous network activity, but this is not so important for this argument. We do not know what these neuron cultures feel or think, what they enjoy, and ethically more relevant, how they suffer. The amount of psychological suffering experienced by the neuron culture could potentially be the same or larger than that experienced by a live rat. Finally, given that computer-based robots are much more efficient and that experiments with these generally do not involve any animals being harmed, one could argue that, if the ethical goal is to reduce the suffering of animals, we should only experiment with computer-based robots.

In Warwick's view, one goal of MEA-based robotics should be to build an oversized brain of human neurons (exceeding 100,000,000,000 neurons). This brain could turn out to be superior to the brains of regular human beings. However, this seems to rule out some of the motivation for the research itself. It seems ruled out, at least from the perspective of rights-based ethics or duty ethics, to perform experiments on such a brain involving giving or predisposing it for diseases such as Alzheimers one of Warwicks arguments for the technology. Given that this being's brain is superior to ours it seems reasonable that we should give it at least the same rights as we have ourselves. This imposes a duty on us not to intentionally harm the being and to not use it as a means to obtain our own goals. From a consequentialist point of view, there might be a better chance of justifying such experiments, given that the benefits of curing a disease such as Alzheimer's would be immense. However, should one not consider the possibility that the utility of such a superior brain is so great as to outweigh the utility of even curing Alzheimer's? In this case, the experiment would not even have ethically justifiable consequences. Warwick also does not seem to acknowledge that there might be an ethical problem with experimenting with human neurons just because they are human and because our brain is intrinsically related to our personhood and dignity.

> ...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.

Warwick (2010)
To many of us the thought of bringing back the dead in this way may sound more like the promise of a zombie movie than of happy memories. More importantly, the dignity of the deceased and of what is human in general is not taken into account here. Using human brains for household robots seems akin to using human skin for lamp shades as was done in Nazi Germany. Is there not something fundamentally wrong in reducing what was once a brain of a real person to the performance of household functions, such as vacuuming the house or making dinner? There are at least some ethical questions that would need to be addressed much more thoroughly before conducting research on MEA-based robots with human neurons.

6 The prospects of MEA-based robotics

I have shown that the current state of MEA-based robotics is not sufficiently methodologically grounded and that more ethical reflection is needed. At the moment, MEA-based robots function poorly and they most likely have no intentionality. They are black boxes on wheels. However, in the long run, MEA-based robotics might still provide fruitful insights and practical applications. It is doubtful whether this will occur before we have a better understanding of the principles of neuron cultures. Given the problems investigated in this paper, this understanding is not likely to come from MEA-based robotics alone, but rather from other areas of neuroscience and neuromechanics. Pending such an understanding, researchers within MEA-based robotics ought to be slightly more modest in stating their results. Overstating results may not itself constitute bad science and might be done for a variety of pragmatic reasons, e.g., attracting media attention and helping obtain funding in the short run. However, in the long run, neither the scientific community nor the general public will benefit from such poor communication. Poor communication might lead us to believe that we have already obtained what we have not and thus delay actual scientific progress towards reaching our goals.

References

Bakkum DJ, Shkolnik AC, Ben-Ary G, Gamblen P, DeMarse TB, Potter SM (2004) Removing some a from ai: embodied cultured networks. In: Embodied artificial intelligence, Springer, pp 130–145

Bialek W, Rieke F, Van Steveninck RdR, Warland D (1991) Reading a neural code. Science 252(5014):1854–1857

Braitenberg V (1986) Vehicles: Experiments in synthetic psychology. MIT press

Chen G, Ying H (1997) Bibo stability of nonlinear fuzzy pi control systems. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology 5(3):245–256

Cowan N, Ankarali M, Dyhr J, Madhav M, Roth E, Sefati E, Sponberg S, Stamper S, Fortune E, Daniel T (2014) Feedback control as a framework for understanding tradeoffs in biology. Integr Comp Biol 54:223–237

Cozzi L, Dangelo P, Chiappalone M, Ide A, Novellino A, Martinoia S, Sanguineti V (2005) Coding and decoding of information in a bi-directional neural interface. Neurocomputing 65:783–792

DeMarse TB, Wagenaar DA, Blau AW, Potter SM (2001) The neurally controlled animat: biological brains acting with simulated bodies. Autonomous robots 11(3):305–310

Dennett D (1984) Elbow Room. Clarendon Press

Dickinson MH, Farley CT, Full RJ, Koehl MAR, Kram R, Lehman S (2000) How animals move: an integrative view. Science 288:100–106

Downes J, Hammond M, Xydas D, Spencer M, Becerra V, Warwick K, Nasuto S, Whalley B (2012) Emergence of a small-world functional network in cultured neurons. PLoS Computational Biology 8:1–17

Ferrández JM, Lorente V, de la Paz F, Fernández E (2013) Training biological neural cultures: Towards hebbian learning. Neurocomputing 114:3–8

Gabbiani F, Koch C (1996) Coding of time-varying signals in spike trains of integrate-and-fire neurons with random threshold. Neural Computation 8(1):44–66

Maeda E, Robinson H, Kawana A (1995) The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. The Journal of neuroscience 15(10):6834–6845

Martinoia S, Sanguineti V, Cozzi L, Berdondini L, Van Pelt J, Tomas J, Le Masson G, Davide F (2004) Towards an embodied in vitro electrophysiology: the neurobit project. Neurocomputing 58:1065–1072

Massobrio P, Tessadori J, Chiappalone M, Ghirardi M (2015) In vitro studies of neuronal networks and synaptic plasticity in invertebrates and in mammals using multielectrode arrays. Neural plasticity 2015

Nishikawa K, Biewener A, Aerts P, Ahn A, Chiel H, Daley M, Daniel T, Full R, Hale M, Hedrick T, Lappin A, Nichols T, Quinn R, Satterlie R, Szymik B (2007) Neuromechanics: an integrative approach for understanding motor control. Integr Comp Biol 47:16–54

Novellino A, D'angelo P, Cozzi L, Chiappalone M, Sanguineti V, Martinoia S (2007) Connecting neurons to a mobile robot: an in vitro bidirectional neural interface. Computational Intelligence and Neuroscience 2007

Pearl J (2009) Causality, 2nd edn. Cambridge university press

Pereda AE (2014) Electrical synapses and their functional interactions with chemical synapses. Nature Reviews Neuroscience 15(4):250–263

Pine J (1980) Recording action potentials from cultured neurons with extracellular microcircuit electrodes. Journal of neuroscience methods 2(1):19–31

Reger B, Fleming K, Sanguineti V, Alford SS, Mussa-Ivaldi F (2000) Connecting brains to robots: An artificial body for studying the computational properties of neural tissues. Artificial Life 6:307–324

Roth E, Sponberg S, Cowan N (2014) A comparative approach to closed-loop computation. Current Opinion in Neurobiology 25:54–62

Stanley GB (2013) Reading and writing the neural code. Nature neuroscience 16(3):259–263

Steiner A, Redish A (2014) Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nature Neuroscience 17:995–1002

Talwar S, Xu S, Hawley E, Weiss S, Moxon K, Chapin J (2002) Rat navigation guided by remote control. Nature 417:37–38

Tessadori J, Bisio M, Martinoia S, Chiappalone M (2012) Modular neuronal assemblies embodied in a closed-loop environment: toward future integration of brains and machines. Frontiers in Neural Circuits 6:1–16

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(1):61–66

Van Staveren G, Buitenweg J, Marani E, Rutten W (2005) The effect of training of cultured neuronal networks, can they learn? In: Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on, IEEE, pp 328–331

Wagenaar DA, Pine J, Potter SM (2006) Searching for plasticity in dissociated cortical cultures on multi-electrode arrays. Journal of Negative Results in Biomedicine 5:16

Warwick K (2010) Implications and consequences of robots with biological brains. Ethics and information technology 12(3):223–234

Warwick K (2012) Robots with biological brains. In: Lina P, Abney K, Bekey GA (eds) Robot Ethics: The Ethical and Social Implications of Robotics, MIT Press, pp 317–332

Warwick K, Nasuto S, Becerra V, Whalley B (2010a) Experiments with an in-vitro robot brain. In: Cai Y (ed) Computing with Instinct: Rediscovering Artificial Intelligence, Lecture Notes in Artificial Intelligence, vol 5897, Springer, pp 1–15

Warwick K, Xydas D, Nasuto S, Becerra V, Hammond M, Downes J, Marshall S, Whalley B (2010b) Controlling a mobile robot with a biological brain. Defence Science Journal 60:5–14

Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, et al (2015) Cell types in the mouse cortex and hippocampus revealed by single-cell rna-seq. Science 347(6226):1138–1142