Cyborgs and moral identity

G Gillett

Neuroscience and technological medicine in general increasingly faces us with the imminent reality of cyborgs—integrated part human and part machine complexes. If my brain functions in a way that is supported by and exploits intelligent technology both external and implantable, then how should I be treated and what is my moral status—am I a machine or am I a person? I explore a number of scenarios where the balance between human and humanoid machine shifts, and ask questions about the moral status of the individuals concerned. The position taken is very much in accordance with the Aristotelian idea that our moral behaviour is of a piece with our social and personal skills and forms a reactive and reflective component of those skills.

Neuroscience and technological medicine in general increasingly faces us with the imminent reality of cyborgs—part human and part machine complexes which function as a whole. Indeed the insertion of a shunt for hydrocephalus is one relatively simple example of the creation of such a hybrid. However, the new and revolutionary developments using intelligent componentry to simulate human vision and to link computer circuitry with voice and interactive technology raise the debate to a new level. If my brain functions in a way that is supported by and exploits intelligent technology both external and implantable, then the prospect of cyborgs takes on ethical significance, raising questions about what is a person and how one should treat a partly artificial being, particularly when that artificiality touches the centre of the creature’s being.

Some fanciful cases

H: Head Injury and neurorehabilitation

Bjorn and Anna have a three year old son, Hansie, with a severe brain injury who, it is predicted, will die. However, they receive new hope when a specialist in “neuroreconstruction” tells them of a technique in which a combination of enzymes, growth factors, and microelectrical stimulation guided by a computer driven three dimensional graphics program could restore Hansie’s brain to a potentially functioning state. The process takes about six months because the neuroreconstructive activity takes a similar amount of time to fetal growth and differentiation of the brain. Hansie will lose all his memories, cannot be guaranteed to have the same temperament as he had before the accident, and will require re-education as if he were a newborn baby (he will more or less catch up with his peers about three or four years after his reconstruction). Bjorn and Anna recognise that they are taking on an enormous task but are so overjoyed at the prospect of not losing Hansie and, indeed, having him back in a way that they hope will turn out well in the
end, that they are more than willing to go ahead.

I will leave aside, for the sake of argument, the ethical concerns about innovative treatment and human experimentation raised by this case and focus on the fact that the procedure contemplated is almost the same as having a new and different child. In particular the “new brain” is moulded by using sophisticated technology under the guidance of experts who are going to dictate to some extent the way that Hansie’s cognitive and personality characteristics are formed. Should they be allowed to think of this child as still being Hansie and does our ethical attitude to what they are doing depend on how we answer that question? I would say that they can and that it does not even though some theories of personal identity would render both of these stands problematic. The justification will emerge below.

V: Vision restoration through cybernetics

It is currently possible to provide blind patients with surrogate vision sufficient to detect things in their environment by using a combination of video camera technology and an array of electrodes delivering impulses to the patient. Let us imagine that the technology reaches the point where the electrodes are directly onlaid so that they stimulate the brain and the person concerned then uses the information in the patterns of excitation that are set up to exploit their visual environment.

If such technology was successful then, I think, we would have little moral concern even though a cybernetic brain was playing an important part in a human individual’s adaptation to their environment and, presumably, influencing that person’s thoughts and feelings about the world to some extent. Our intuitions here seem to be based on the fact that the core of the person, that which defines his or her identity and relationships remains untouched by this technology which functions very much like an add on or prosthesis used in service of a life with its own holistic integrity.

HC: Extensive brain injury and replacement with micronetworks

Henri-Charles’s neuroreconstruction is very similar to Hansie’s, with one difference: it is facilitated by and incorporates a set of relatively modular microchips with the capacity to respond to biological signals generated by cell membranes and neural cell columns. The modular chips subserve some of the well established functional connections that are known to occur in the human brain to do with sensorimotor interaction in perceptual cycles and some stereotypical responses such as those involved in affect programs, and some of the assemblies created replace damaged neural tissue.

This case is exactly like the Bjorn and Anna case and is also science fictional. In this case the result is a truly cybernetic brain in which a number of centralised and widely connected functions (such as affect programs) are designed in ways which exemplify general patterns of connectivity. These are then integrated into the cerebral neurocognitive system in an idiosyncratic way by the individual concerned, which is similar to what happens during neurorehabilitation. Our intuitions are thrown, however, by the fact that some of the human reactions and responses that Henri-Charles will come to exhibit may be cold, calculating, and robotic in ways that disengage him from the moral community at important points; we can almost hear the humming of the circuits which plan and execute his responses. The thought of his alien rather than fully flesh and blood nature tends to capture our thinking and make us wary of him: but such prejudice will not do for moral thinking and we need something to take us further than conjecture and somewhat unconstrained imagination.

A: Anencephaly with brain simulation

Bryn and Lilith are “older” parents with Lilith not expecting ever to have another child. Unfortunately the child who is to be born to them, whom they had intended to call Arlo, has been shown by intrauterine scans to be almost anencephalic and his brain, though covered by intact skull and skin, is only a series of membranes containing spinal fluid draped over a primitive brain or brainstem like structure comprising rudimentary neural assemblies and circuits connected to the top of the spinal cord. They are approached by Dr F whose neurosimulation team offers to use cells taken from (to be) Arlo’s umbilical blood and primitive neural tissue to interface with a cybernetic assembly reproducing much of the complexity of a neonatal brain. They have used this technology in some higher animal infants with severe developmental abnormalities of the brain and these creatures have taken on characteristics virtually indistinguishable from those of normal individuals of their species.

The case of Arlo is an entirely fanciful projection (to a certain limit) of possibilities raised by visual simulation and Henri-Charles and it confronts us with the cyborg issue in a stark way. It asks the question: “What is it to be human” in a way that does not seem so pressing for the cognitively similar case of Hansie. Some might regard this as so fanciful that they cannot form any clear moral intuitions but the sources of unease are laid out in the case of Henri Charles and so I will offer one more thought experiment on this topic to enrich the grounding of our intuitive reactions.

P: Personality upgrade with micronetworks

The following story extends our intuitions about cyborgs (in a similar way to the film Bicentennial Man).

Peggy and Bob are married but, sadly, not happily so. They began married life as a fairly typical young married couple but then Peggy fell prone to episodes of depression, spending more and more time at home, not working, and lamenting her inability to have children. She has tried antidepressants and psychotherapy but no progress has been made and she has increasingly taken to her bed, neglecting herself, her marriage, and any hope of a career. Bob is desperate as he is increasingly fearful of neglecting herself, her marriage, and any hope of a career. He goes to the centre and is shown a recently released companion/carer model mainly used for caring for older folk but which can be used also with young disabled people. He purchases an android, calls it Andrea, and takes it home.

Andrea is marvellous. She brings tea in bed, does the housework, and is a companion/carer model mainly used for caring for older folk but which can be used also with young disabled people. He purchases an android, calls it Andrea, and takes it home.
and so on. Quite soon he also finds that he thinks of Andrea as a person.

After an extended business trip he comes home to a Peggy who has made significant gains in her psychological functioning and he finds that she has had some treatment in the revolutionary neuropsychiatric clinic recently opened in their city. She is the Peggy he married—active, positive in her attitude, and relaxed about life. After his next trip he finds she has taken up watercolour painting and joined a reading group. In fact, Peggy has improved so much that when the time comes to renew Andrea’s lease, she is returned to “Cybo-help”.

Bob and Peggy miss Andrea but are very much happier than they have been for years until one night. Bob is stroking Peggy’s hair and feels a row of letters embossed in her scalp behind her hairline.

Should he worry?

ETHICS AND THE HUMAN: INTUITION AND REFLECTION

The spectre of robots with human attributes has always been the stuff of science fiction but advances in implantable microchip technology and prosthetic devices which can interface smoothly with the human brain make our moral concerns a lot more pressing than they used to be. In the cases I have described we find ourselves left at tantalising points with one or two prima facie (and very broad brush) pegs to hang our thinking on.

(i) We are less concerned when the cybernetic components of the person seem peripheral or somewhat incidental to their psychological identity or character.

(ii) We are more concerned where a non-human mode of relationship and reaction or response to others may affect a person at a very deep and pervasive level.

A moment’s reflection, however, on our intuitive idea that a machine, no matter how human like, is not a candidate for moral properties, reveals that it is an idea that is hard to justify. It is even harder to emotionally sustain after a film like Bicentennial Man or AI, where we see the world through the narrative eyes of such a character. When do we that we are drawn to identify with them and realise that they strike many of the same balances that we do between expedience, attachment, sensitivity, and so on.

It is on the basis of just such things that we have moral attitudes to human beings such that we regard it as wrong to do certain things to them and feel that it is good to treat them in certain ways and show them certain kinds of consideration. The peculiarities of those attitudes are internally related to exactly the properties of relatedness, vulnerability, and so on that arise between people. Our attitudes to human beings and their interests are distinguished by these features from the sense of right and wrong that applies quite straightforwardly to machines. It is wrong to put diesel fuel in my car and we understand what is meant when someone says that a machine is “happy”. These evaluations are related to functions for which the artefacts in question have been designed and they differ from evaluations of right and wrong in human cases in part because there is no agenda belonging to the individual or robust subjectivity to provide a point of application for our sympathies in such cases. The relationship is entirely functional and instrumental and there is no admixture of organic intertwining of feelings and first person narratives in our dealings with machines. We can make this thought vivid by imagining a practical joke that would work in a stage one philosophy class. Imagine a group of first year philosophy students turning up to their tutorial and being confronted by a tape recorder. When they are seated the tape recorder “says” in a mellow female voice “I think, therefore I am”. The group laugh. The tape recorder goes on, now sounding a little desperate, “Don’t laugh at me!”

The group look quizzically at each other.

The tape recorder, in a pleading tone, says: “Ok, Ok, I know why you are looking at each other like that, but just believe me I got morphed into this form and I am desperate for human interaction”.

As one of the group points at the recorder and says, “What’s with this thing”, the recorder says, “Please, don’t turn me off, I live for these times!”

Now this tape recorder may sound and appear to cogitate and interact with us exactly like a creature that we believe to be a soul (a moral existent among us) but we do not believe that it is one. We do not think that the simulated narrative and subjectivity are anything more because the parameters of function are explicable on a much more parsimonious basis. We are similarly not fooled into entertaining the thought that cyberpets who need petting and talking to or they “die”, are anything but clever simulations on hand held computers even though they are designed to respond somewhat like real pets. Wittgenstein makes a telling remark about non-human souls.

Look at a stone and imagine it having sensations. One says to oneself: how could one so much as get the idea of ascribing a sensation to a thing? And now look at a wriggling fly and at once these difficulties vanish and pain seems to get a foothold here, where before everything was, so to speak, too smooth for it.²

He is surely right in that we cannot get our heads around the idea that suffering is or could be manifest in the sense that it matters morally (rather than a gearbox grumbling and us wincing with the metaphorical “pain” of it, or a cyberpet showing a “sad” face) when its “subject” is not a living flesh and blood creature. But even (animal) pain or pleasure in and of themselves seem a little “thin” to form the basis for substantive moral attitudes (though we might not feel morally comfortable with the treatment of flies by wanton boys). But the mortality and sufferings of organisms such as tapeworms and insects are traditionally of no account alongside that of the cats and dogs we treat by eradicating them, so we have to do a bit more work to get to the reason why it is that the grimaces, writhings, and struggles of a tapeworm or an ec (as for an android) “do not cut it” for us in terms of moral engagement. The thought is that these things are much too limited to be ends in themselves or have subjectivities of their own and can only supplement the stories they figure in which have a narrative point of subjectivity located elsewhere. We are interested in the subjectivity at the centre of the story and need that posit to be credible in order for the being to engage our moral concern.

We can refine our intuitions further by considering the kind of goodness and badness we recognise in our treatment of animals (even when we find their reactions to what we do quite inscrutable). For instance, it seems undeniably good to provide suitable conditions for an orang utan, for whom we may think that there is an absolute sense of good linked to having available a rain forest habitat in which its nature can be expressed. This natural kind of good is regarded as a kind sufficient for generating a robust conception of human goodness by some contemporary writers.⁴⁵ Whether or not that is true, we can understand the idea that such attitudes concern our being and have genuinely ethical subject matter.

If, however, artefacts cannot get into the moral domain through any resemblance they may have to organisms, despite their cognitive engagement with us, what is doing...
the ‘soul work’ here? Somewhere in the scale of organisms we
got to the kind of pleasure and pain and other responses
that move the creature concerned into the moral domain for
us. How do we begin to understand that point? I have
gestured at clues to be found in the idea of a narrative.

“Hal” from 2001, a Space Odyssey comes across as a moral
agent because of the way he interacts with human beings and
establishes a cognitive/intellectual rapport with them (as
distinct from the more “earthly” organic rapport one
enjoys with a living creature). We also get the sense that he
has a history that he is living through. Descartes’s tape
recorder goes some way towards doing the same as Hal when
he fills in his background for us and a cyberpet may simulate
the sense of life at stake were it to “remember” its history of
relationship with its owner, convey disappointments, recriminations, joy at being awakened by its friend, and so on. In
each case it is understandable that we treat the other to
whom we relate as a moral being because some aspects of our
normal ways of interrelating with moral agents are in play
and we fill in the rest (in the way human beings are
cognitively prone to do). Most of the time this “filling in” is,
of course, entirely appropriate because the being before us
cognitively manifests at any time only a subset of a complex
constellation of natural abilities and capacities (including
those which give rise to moral reactions, responses, and
exchanges that justify and reciprocate our regard for them)
that are involved in living out a life story among others. The
filling in or ascription of a moral state of being to an
interlocutor then draws on a history of similar encounters
to generate its own momentum and cumulative context (where
it is reciprocated). Thus we are souls whose lives intersect in
the moral realm.

WHAT IS A HUMAN SOUL?
The soul, for an Aristotelian thinker, is a complex and
integrated whole emergent from a set of biological and
psychological functions that characterise a creature of a given
type. Aristotle discusses this in De Anima and also in
Hume’s An Essay on Human Understanding, and this view is neatly summarised by
Locke in his Essay on Human Understanding. A human being
has vegetative, animal, intellectual, and social or political
functions which combine to give individual form to the life a
person lives among his or her fellows. We could say that the
narrative continuity of the human soul is based on the
temporally extended flesh and blood existence of critters like
us. How do we begin to understand that point? I have
gestured at clues to be found in the idea of a narrative.

What is it then that we denote by the word ‘soul’? What are
the related things: (i) reactive responses to those things; and (ii)
representing to ourselves the form of a thing and factoring
that into our reasoning about our treatment of it. An
Aristotelian recognises that a particular form of an object,
say a chair, realised in bronze, is not the same as one similar
in every other respect but realised in wood. In the case of a
soul we can imagine a holism about the habitus (including
responsivity, energy, and vulnerabilities among other things)
of a human being that is uniquely grounded in human flesh
and blood and cannot be reproduced in silicon, any other
fabricated material, or even different flesh. We could argue
that the human being combines uniquely feeling, fleshly
contingency, and intellect (broadly construed) in a holistic
way that defies reductive analysis so that even this three
point delineation is somewhat crude and misleading. We
would then need to make plausible the idea that only a
human being with the holistic nature that implies life in
our kind of subjectivity (a thought that is not that
inaccessible)—could you come back as a being differently
gendered from the one you are?

It now begins to seem as if it is the total form that is
revealed in a lived life story that gives a being the identity
which matters morally and that identity, in the sense we
respond to it in our moral thinking, is somewhat indifferent
to the material of which the being is made except in so far as
that material affects the relevant lived experience (which is
not independent of how we react and respond). Thus, for
instance, if a person could not respond to me in a
characteristically human way because the megabytes that
control them are unresponsive to the hormones and
biochemical changes that influence human reactions
and feelings, then the contribution of the artificial aspects of
the cyborg (to the whole being with whom we are
interacting) has affected the being of the (candidate) person
in a morally relevant way and we might find that moral
responses and judgments differing in important ways from
our responses to our fellows are appropriate. It seems to be
just this intuition that Shylock appeals to when he makes his
famous “Hath a Jew not flesh and blood” speech to highlight
his congruence of reactions with those judging him. We can
imagine this not working with Arlo and the transformed
Peggy but ringing true in the case of Hansie. Our intuitions
about Henri-Charles are I think sufficiently shifting for us to
have to “see how it goes”.

To this holistic appreciation of the fact that one is in
the presence of a critter like oneself one brings a certain reflective
or perceptual equilibrium, involving both intuitions and a
rational analysis of the facts surrounding relevant encounters
and their characteristics. In the end, however, one judges
according to the responses ones finds evoked in oneself and
their characteristics over time, and to reflection, in much the
way that the Aristotelians claim. Faced with this strange
moral fruit we, “suck it and see”. I cannot, however, see that
the tissue of which one is composed has, apart from its effect
on the individual’s ability to participate in our forms of life,
any more moral relevance than the colour of one’s skin in our
relationships to that being. Is the being before me able to feel
pain? Does the being before me develop attachments and
make an appeal to me? Does the being before me have a story
in which moral participation features? What we ought to do
seems to be the result of the myriad normative demands and
imperatives that we are inculcated into in our daily
interactions with those who shape us into the individuals we
are.

Others shape me by imparting skills, for instance of self
directed and self formation, that enable me to take some
control over the directions in which my life goes. I obey those
“oughts” and “shoulds” or reject them and grow as an
individual in the shared soil of my culture and my land. Thus
we are beings who have inscribed deeply into us our ways of
being that include belongings, oughts, and desires, inspirations
and aspirations, skills and styles of moving, sources of
energy and hesitation, taboos and permissions, and so on.
Therefore what we ought to do is to be true to our nature as beings who live as members of a kingdom of ends able to recognise, take account of, and respond to each others' subjectivities as they are revealed in lived experience when we interact with each other and tell our stories.\textsuperscript{11} For that reason, in any imaginable case, I think we \textit{ought} to react on the basis of a sum, albeit complex, dynamic, and impossible to reduce to formulations, of the mutual participation in language games where morality is relevant. On the basis of that complex engagement in a many faceted discourse, our conception (metaphysical if you like) of \textit{what a human being is} is derived from the beings with whom we share these formative and sustaining interactions.

THE CASES REVISITED

The cyborg cases can now be resolved (or at least approached with resolve). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes).

The creature concerned ought then to be treated as such an acquaintance would treat them. Others and a sense of life. The cyborg cases can now be resolved (or at least approached with resolve). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes). In each case we evaluate somewhat informally the extent to which we have a genuine human being among us on the basis of myriad cues that are manifest in our forms of life. These cues and clues are drawn for us by our dealings with the persons in question and reactions (or reactive attitudes)\textsuperscript{12} are evoked in us depending on our dispositions with the persons in question and reactions (or reactive attitudes).

We are left with the problem of Peggy, a problem bedevilled by the question: “What has happened to the real Peggy?” This problem keeps company with the thought that there is an android intelligence in Peggy who may be like a second personality—knowing but not known and ultimately able to perform a coup aided by a scheming cyborg, Andrea. This problem is an epistemic, metaphysical, psychological, and forensic conundrum but I would argue that the epistemic virtues needed to gather the data relevant to the metaphysical question cannot be exercised in the absence of the right moral attitudes.

It therefore seems to me that a cyborg is, on the present account, as human as his or her life among us indicates to those who approach the encounter with an openness to others and a sense of life. The creature concerned ought then to be treated as such an acquaintance would treat them.

REFERENCES