Postdigital Science and Education
https://doi.org/10.1007/s42438-019-00095-2

ORIGINAL ARTICLES

Ethical Conundrums and Virtual Humans

John Reader[1] · Maggi Savin-Baden[2]

Published online: 19 December 2019

Abstract
This paper explores ethical conundrums and virtual humans through building upon a post-Kantian framework, and one emerging from what is known as New Materialism. It begins by presenting the recent research and literature on virtual humans and suggesting that the central ethical conundrums that need to be examined are those of agency and values. The paper then argues that a combination of Luciano Floridi's approach and one developed from New Materialism, namely modest ethics, offers a means of engaging with the ethical conundrums of virtual humans. It is argued that as yet there is little evidence for a democratic design process for virtual humans nor is there evidence about the possible impact virtual humans may have on a postdigital society. The paper concludes by suggesting that there need to be more open processes for debate which bring to light the values that are being built into these profound developments by the experts and focuses on using a modest ethics approach.

Keywords Virtual humans · New materialism · Modest ethics · Agency · Artificial intelligence

Introduction

One of the concerns articulated in the paper and supported by work such as that of Bernard Stiegler (2016) is that the current economic and political structures have determined the development and deployment of virtual humans to serve certain vested interests. This raises the question of whether and how there could be alternatives based on earlier interventions and a different understanding of how humans interact with the non-human and the digital. A combination of Luciano Floridi's approach and one

✉ Maggi Savin-Baden
m.savinbaden@worc.ac.uk

John Reader
DrJohnReader@hotmail.co.uk

[1] Honorary Senior Lecturer with the School of Education, University of Worcester, Worcester, UK

[2] Professor of Higher Education Research, School of Education, University of Worcester, Worcester, UK

developed from New Materialism offers a means engaging with the ethical conundrums of virtual humans. Floridi (2011, 2015) sets out to develop a framework encompassing the whole of reality which incorporates digital technologies as part of an expanded understanding of human autonomy.

New Materialism emerged as a term during the late 1990s, primarily in the context of interpretations of the philosophy of Gilles Deleuze and was undertaken as a response to a with the crude reductionist forms of atomic materialism. New Materialism affirms such diverse processes as energy transformation and cellular reproduction but rejects the idea of a hierarchical scale of values. Two main strands have developed; the first associated with thinkers such as Levi Bryant, Jane Bennett, Rosi Braidotti (2013), and Manual DeLanda (2002), and the second with a more explicitly scientific interpretation linking back to the work of Alfred North Whitehead and represented by Isabelle Stengers (2011), Diana Coole and Samantha Frost (2010).

This paper the emphasizes the ethical implications of this approach and what we intend to describe as a modest ethics. Two of the central concerns of this article are firstly that there is currently no democratic process for the design of virtual humans. Secondly, that as yet little is known about what the impact and unintended consequences of creating virtual humans will be. Hence there are concerns both about the process by which future developments will be shaped and how, for instance, non-experts might be involved in this, and also the consequences of such non-involvement on the shaping of those developments. What values and interests will drive future developments if these are left to purely commercial and corporate forces?

Virtual Humans in Context

One of the challenges of discussing virtual humans is that everyone appears to have a different idea about what counts as a virtual human. It is evident from the literature (reviewed in-depth by Burden and Savin-Baden 2019) that the term virtual humans tends to be used as an overarching one that includes other terms such as Chatbots, Conversational Agents and Pedagogical Agents. Virtual humans are seen as human-like characters on a computer screen or agents with embodied life-like behaviours which may include speech, emotions, locomotion, gestures, and movements of the head, eyes, or other parts of the body. Discussions around virtual humans often bring in the concept of machine learning and, in particular, neural networks. Machine learning is defined here as a computer system that can learn to make decisions based on the examination of past inputs and results, so that its future decisions optimize some parameter – such as perhaps recognizing faces in photographs. Whilst a machine learning system could well be *part* of a virtual human, it is certainly insufficient to be a complete one. If the term 'machine learning' is actually used to mean a neural network-based system, then it is probably not a necessary one either.

The result is the temptation to choose a behaviouristic definition of a virtual human, typified perhaps by Turing's original Imitation Game (Turing 1950). Behaviourist definitions have, though, been frequently challenged (for example Searle 2014), along with Chalmers' (1996) arguments about zombies being portrayed in ways that might be logically but metaphysically impossible. So perhaps useful working definitions, following (Burden and Savin-Baden 2019), are:

- Virtual Humanoids are simple virtual humans reflect some of the behaviour, emotion, thinking, autonomy and interaction of a physical human.
- Virtual Humans are software programmes which present as human and which may have behaviour, emotion, and interaction modelled on physical human capabilities.
- Virtual Sapiens are sophisticated virtual humans which achieve similar levels of presentation, behaviour, emotion, interaction, self-awareness and internal narrative to a physical human.

Harari, in *Homo Deus* (2015), explores the projects that will shape the twenty-first century. Much of what he argues is disturbing, such as the idea that human nature will be transformed as intelligence becomes uncoupled from consciousness. Google and Amazon, among others, can process our behaviour to know what we want before we know it ourselves, and as Harari points out, somewhat worryingly, governments find it almost impossible to keep up with the pace of technological change. Harari challenges us to consider whether indeed there is a next stage of evolution and asks fundamental questions such as: where do we go from here? Perhaps what it means to be human is bound up with understanding of consciousness or the idea of having a soul. This is something Harari explores but he seems to mix up religion and values by using the label ‘religion’ for any system that organizes people, such as humanism, liberalism and communism.

Harari argues that ‘it may not be wrong to call the belief in economic growth a religion, because it now purports to solve many if not most of our ethical dilemmas’ (Harari 2015: 2017). However, what is not clear in discussion about the relationship between humans and virtual humans is whether technology is shaping our behaviour and our relationships. Similarly, Reader asks: What exactly is the relationship between humans and technology and could it be the case that any distance between has been so eroded that any sort of critical perspective is lost in the process? How, if at all, can we absent ourselves when the impacts of digital technology are so pervasive and invasive? (Savin-Baden and Reader 2018). Whilst descriptions of virtual humans are useful to the debate, it is important to review the types of ethics that relate to the issue of virtual humans.

Ethical Conundrums

Malle (2016) argues that robot ethics needs to examine questions about how humans should design, deploy, and treat robots. He suggests that two issues need to be explored, both of which apply to virtual humans, first ethical questions about how humans should design, deploy, and treat robots, and secondly questions about what moral capacities a robot should have. The history of robot ethics is largely seen as stemming from Isaac Asimov’s famous Three Universal Laws of Robotics in 1950 (Asimov 1950). These were based on the idea of robots being ethical themselves, but the field of robotics ethics has become considerably more complex since then with the focus being on researching robot ethics and ethical robots. The former is concerned with ethical use of autonomous systems, while the latter is concerned with how autonomous systems can themselves be ethical (Winfield et al. 2019). Currently, robot ethics features such topics as ethical design and implementation as well as the considerations of robot rights.

Hern (2017) reported that The European Parliament has urged the drafting of a set of regulations to govern the use and creation of robots and Artificial Intelligence (AI). However, machine morality explores issues about what moral capacities a robot should have and how these might be implemented. It also includes issues such as moral agency justification for lethal military robots the use of mathematical proofs for moral reasoning (Malle 2016). There are a range of debates on whether robots and virtual humans can have both emotions and empathy. For example, Prinz (2011) suggests that empathy is not needed for moral judgement whereas moral judgement does require emotion. However, Docherty (2016) argues that robots and robot weapons lack both emotions and empathy and therefore cannot understand the value of human life.

Engaging with virtual humans offers people opportunities to connect with something emotionally and feel supported, even loved, without the need to reciprocate. Although this kind of inter-relationship is becoming more common (Levy 2008) it introduces challenging questions about what it means to be human. Malle argues that.

> any robot that collaborates with, supports, or cares for humans—in short, a social robot—poses serious ethical challenges to the human design and deployment of such robots, and one of the most important challenges is to create a level of moral competence in these robots that is adequate to the application at hand (Malle 2016: 244).

Borenstein and Arkin (2016) argue that the main issues that need to be considered are affective attachment, responsiveness and the extent and duration of the proximity between the human and the robot. Emotional connection has also been found to be one of the strongest determinants of a user's experience, triggering unconscious responses to a system, environment or interface. For example, Reeves and Nass (1996) argue that humans do relate to virtual humans in similar ways to humans, but that voice and affect are central to agency attribution. Recent work in this area would support this (Savin-Baden et al. 2013), as it found that emotional connection with pedagogical agents is intrinsic to the user's sense of trust and therefore likely to affect levels of truthfulness and engagement. The implications of this study are that truthfulness, personalisation and emotional engagement are all vital components in using pedagogical agents to enhance online learning. The study also indicates the importance of the need for behavioural (as opposed to photorealistic) authenticity of the virtual human, in terms of it being effective in influencing humans. Thus, an ethical concern is the extent to which a robot affects a human's autonomy, for example, whether those people who have trouble forming relationships will develop an over-dependent relationship with a robot which reduces their ability to act as an autonomous human being.

Virtual humans are designed using human decision criteria and therefore ethical behaviour need to be 'designed-in-to' virtual humans. However, in the context of the range of virtual humans that are being developed, designing appropriate and effective ethics standards are complex and far reaching. For example, if autonomous combat robots or AI are deployed by the military whose fault is it if they mistakenly attack or fail to distinguish correct targets? Riek and Howard (2014) suggest that design considerations should include:

- reasonable transparency in the programming of robotic systems,
- predictability in robotic behaviour,
- trustworthy system design principles across hardware and software design, and
- opt-out mechanisms.

What is clear is that there is little legal guidance, and, in many instances, current ethical guidelines supplied by governments, universities and professional bodies fail to deal with the changing challenges of ethics and virtual humans.

Much of the debate that occurs in the public sphere about the place of virtual humans in society focuses on whether they will take over jobs and lives, as discussed often at the UK's House of Lords Select Committee on Artificial Intelligence (2018). However, the issues are much more complex than this. For example, questions about the use of virtual humans for the social good introduce questions about what counts as the social good and who decides? Can or should virtual humans have rights? Are there different levels of virtual humans that require different levels of ethical stances? Underpinning all of these questions is the need to delineate the values that inform the development, creation and management of virtual humans. This in turn raises questions of agency.

Work about agency in relation to virtual humans has concentrated on attributed agency. Attributed agency is both the sense of agency and the way in which agency is attributed to a human, virtual human or digital immortal. In the case of virtual humans what seems to be evident is that the context affects the sense of agency. For example, Obhi and Hall (2011) found that humans consider face to face shared action with other humans different from human-computer shared actions. The findings indicated that attributed agency tends to be over-ruled when the participant is aware the computer is a co-actor. Studies such as this seem to imply that humans are more likely to attribute agency to other humans but not to virtual humans.

Posthumanism seeks to break down binary distinctions between 'human', 'machine' and 'text' (Hayles 1999, 2012) and between 'nature' and 'culture' thus also rejecting dualisms that are used to define beings as either subject or object. Thus, it is a theory that is used to question the foundational role of 'humanity' and prompts consideration of what it means to be a human subject, and the extent to which this idea is still useful. This has overlaps with Actor Network Theory (Latour 1987) where the arguments centre on the idea that actors may be both human and non-human, thus for example supermarket products and digital devices are seen as actors that have influence, but it not clear how all this relates to the creation of a virtual human.

Questions about the exact nature of human agency and particularly the relationships between the human and the non-human are central to New Materialism. A possible weakness of this approach is that its understanding of what it is to be human detracts from the capacity of humans to be effective agents, particularly at a social or political level. It will be the argument of this paper that some New Materialist interpretations in fact add to rather than subtract from a necessary concept of human agency and provide a more adequate base for political engagement. One needs to be aware though that the New Materialist writers do not share a united view on this issue but represent a spectrum of possibilities. The key question is that of whether and how this recent work retains a sense of agency adequate for social and political engagement, whilst balancing it with a more limited notion of human autonomy and which allows greater scope for the agency of the non-human, notably as in constant and shifting assemblages of both human and non-human.

Drawing upon Latour (1987, 2004), and more closely associated of course with Actor Network Theory, the argument is that there are many actants or agents at work in the world and that humans are only a small fraction of these. Relationships between humans and non-humans and indeed between non-humans, are entangled and enmeshed and have to be understood as such. As the New Materialist authors DeLanda (2010), Bennett (2010) and Bryant (2014) argue, humans and non-humans form part of assemblages that need to be identified and examined example by example. Latour's main concerns are more environmental and represent a need to redefine the traditional distinction between nature and society. If it were as simple as bringing those two into a straightforward relationship, then the ecological crisis which is now upon us would have been averted. Instead, what is required is a new way of understanding the collectives or assemblages which incorporate both nature and society and then to allow the non-human a voice in the process of political debate.

> We are going to show how humans and non-humans, provided that they are no longer in a situation of civil war, can exchange properties, in order to compose in common the raw material of the collective. Whereas the subject-object opposition had the goal of prohibiting any exchange of properties, the human-non-human pairing makes such an exchange not only desirable but necessary. This pairing is what will make it possible to fill up the collective with beings endowed with will, freedom, speech, and real existence. (Latour 2004: 61).

The apparently unlikely outcome that Latour is trying to achieve, is that non-humans should have their own voice, and they can do this through the intermediaries of spokespersons, someone who can speak in their place. To describe these intermediary states Latour (2004: 64) argues 'we can use the notions of translation, betrayal, falsification, invention, synthesis or transposition'. His argument is that this is what scientists in their lab coats are doing most of the time, and that they can and do speak on behalf of those other actants. The lab coats have invented speech prostheses which allow non-humans to participate in the discussions with humans, especially when they are perplexed about the participation of new entities in collective life. Effectively, the barrier between science and politics is broken down by these means. This challenges the traditional distinction between subjects and objects. By restoring both human and non-human to civil life, they can both shed the old garments that marked them as subjects and objects in order to participate collectively in what Latour calls 'the Republic' (2004: 71). However, it could be argued that the division between things and people still remains, and that our gaze, as if we were watching a tennis match, is now turned towards objects, and now towards subjects. Latour wants to use this image in a positive sense though, as rather than referring to two different spheres of activity, the shared goal of engaging in the game requires the attention of one to the other that he is trying to advocate. The human non-human pairing does not refer us to a distribution of the beings of the pluriverse, but rather to an uncertainty, a profound doubt about the nature of action, and indeed a variety of positions that make it possible to define an actor. These are part of the matters of concern in which humans participate with non-humans. Although Latour would not count himself a New Materialist author, the influence of

Deleuze is still apparent as he develops his notion of assemblages of the human and the non-human. DeLanda (2010) also draws on Deleuze's meaning of assemblage:

> What is an assemblage? It is a multiplicity which is made up of heterogeneous terms and which establishes liaisons, relations between them, across ages, sexes and reigns - different natures. Thus, the assemblage's only unity is that of a co-functioning; it is a symbiosis, a 'sympathy'. It is never filiations which are important, but alliances, alloys; these are not successions, lines of descent, but contagions, epidemics, the wind. (Deleuze 1987: 69).

Although this offers a broad picture of how he wants to use the term, it only goes so far in addressing more critical questions such as when does an assemblage count as such and how and where does one draw the line.

Ethics, Virtual Humans and New Materialism

The intention now is to examine two different approaches to ethics and virtual humans. The first is that of Luciano Floridi, and the second is one that some of us are developing from New Materialism but is not intended to be an uncritical appropriation of that. The implication which derives from both approaches is that previous ethical frameworks or approaches are inadequate to address the issues and dilemmas that are now emerging. The Floridi approach (2011, 2015) is a bold and creative one which attempts to build upon existing interpretations, notably that of Kant (1998), but with significant developments. While acknowledging the value of Floridi's approach, the alternative to be presented emerges from a different tradition of continental philosophy and argues for greater discontinuities with earlier ethical approaches. It may be that Floridi's position is better placed to address existing technological developments at the level of governance and regulation, but that to more fully engage with future developments something more fluid and radical is required. In particular, as we have seen, discussions about virtual humans are highly technical and inaccessible to most non-experts, while the potential consequences of these developments are profound in terms of how humans understand themselves. What is being suggested is that this is not an 'either-or' argument but more of a 'both-and', reflecting the modesty and limited claims which are characteristic of this alternative approach. In many ways the idea of modest ethics can be aligned with the idea of the postdigital (for example Jandrić et al. 2018) since both prompt a rupture in our existing theories. The modest ethics approach rejects over-arching ethical frameworks and their limited respect for the empirical complexities of real-world scenarios and the power and agency that such frameworks ascribe to human rationality. Modest ethics embraces a deep interconnection between ethics and materiality, leading to a new kind of worldly ethics, which is more-than-human in its scope.

Floridi's work (2015) is part of a larger project including an earlier book *The Philosophy of Information* (Floridi 2011). Floridi himself agrees that there is much work still to be done in this area hence there is a level at which all of this work is exploratory and tentative. Nevertheless, by engaging with Floridi's theory as a substantive position, it will become clearer where and how the differences with the alternative approach are important. The claims that he makes are quite radical. For instance: 'We are modifying our everyday perspective on the ultimate nature of reality, from a materialist one, in which physical objects and processes play a key role to an informational one' (Floridi 2015: 10). One implication of this is that the right of usage will be perceived to be as at least as important as the right to ownership. This is an ontological claim, one about the nature of reality itself, and one can see from the perspective of current technological developments why this would appear to be a valid direction to take. Floridi talks about the Infosphere and the way in which we will all be living 'onlife' and 'online' in due course, and how those who are prevented from participating in this will be the wrong side of a digital divide. From a political and regulatory perspective this is an important argument, but it does preclude the alternative interpretations of materialism presented by the New Materialisms (Reader 2017). What might these have to offer to the debate?

Floridi suggests that Information Ethics is a macroethics, thus making claims about not just the nature of reality, but the ways in which politics and governance should engage with the issues raised. In order to do this Floridi (2015: 31) presents an argument for what he calls 'levels of abstraction'. This raises a number of contested philosophical positions which, again, are difficult to do justice to in a short article but which do need to be registered. Floridi is making universal claims for his approach, ones that are a development of a Kantian position in that respect.

> Understanding the nature of IE (Information Ethics) ontologically rather than epistemologically, modifies the interpretation of its scope and goals. Not only can an ecological IE gain a global view of the whole life-cycle of information, thus overcoming the limits of other microethical approaches, but it can also claim a role as a macroethics, that is, as an ethics that concerns the whole realm of reality, at an informational level of abstraction'. (Floridi 2015: 27).

Floridi (2015: 33) acknowledges pluralism without descending into what he calls a relativism in which 'anything goes'. There are two concerns here. The first concern is that there is a serious challenge to Floridi's interpretation of relativism from within philosophy. For instance, Latour suggests instead that relativism in itself is not a bad thing and is not to be equated with an 'anything goes' argument as Floridi suggests (Latour 2004: 12). There may be less distance between Latour and Floridi than might at first appear, but this does suggest that Floridi is still operating from within a Kantian framework and one which presents us with a binary that is worth challenging: either universalism or relativism.

The second concern is that the concept of assemblages, especially as deployed by DeLanda (2010), is not so different in its intention of finding ways of interpreting and analysing reality at different levels. However, the advantage of the idea of assemblages is that they are contingent, shifting and dynamic alliances of different elements and components, both human and non-human, and therefore present a more fluid understanding of the relationships between humans and the digital technology. This would allow for the ways in which the technologies shape humans as much as the humans shape the technology, a critical point when it comes to understanding the limits of human agency. Floridi is still wedded to what is a development of the Kantian

(1998) position of the existence of autonomous agents, even though he wants to extend this beyond the human in a way that Kant would never have done.

In contrast to Floridi's universalist approach, the alternative we are building upon and beyond New Materialism is more modest or limited in its claims, thus perhaps better able to acknowledge and take into account different interpretations without reducing them to some form of relativism (Reader and Evans 2019). This does not mean refusing to make any claims, but a greater willingness to work across disciplinary boundaries and to acknowledge the contribution of different traditions, including religious ones.

When Hegel (1821) said that the owl of Minerva only spreads its wings with the falling of the dusk, he could have been referring to all attempts to address the issue of an ethics of the digital. What he probably meant was that philosophy generally comes too late on the scene and only grasps the true nature of an historical period once it is drawing to a close. Is the same inevitably the case when it comes to developing an ethics which seems unable to get ahead of technological developments? Once systems and technologies are already in place, only then is the moment to evaluate and to try to regulate in order to contain the worst and most damaging excesses. Floridi's ethics of information is an attempt to establish some public criteria by which to assess and control these developments at an earlier stage in the process, but does it go far enough?

In the 'both-and' scenario for which we are arguing here, the suggestion is that alternative philosophical sources can be employed to engage these developments in ways that Floridi is unable to envisage. It is important therefore to understand the claims that he is making for his approach. Whereas many standard ethical approaches focus on the agent, Floridi's claim centre on less orthodox frameworks such as medical ethics or bioethics, an equal or greater concern with the patient. However, even these alternative approaches are still biased against the inanimate, intangible, abstract, engineered and artificial entities such as information and communication technologies and any form of virtual human in the future.

We suggest that the concepts from a New Materialist understanding supplemented by ideas from Latour as in our modest ethics, achieve the same objectives but without having to make universal claims. The non-human in relation with the human but as part of a flat ontology rather than a hierarchical one provides a more effective means of interpreting the engagements with the inanimate and artificial. The question to be addressed is surely that of how the non-human in the form of virtual humans might change the nature of these relationships and require a different conceptuality. Later in *The Ethics of Information* Floridi (2011) does address this issue, arguing that all informational entities have an intrinsic moral value, although this may be minimal and overridable, and that therefore they will qualify as moral patients demanding respect (2015: 110). Also, 'artificial informational entities, insofar as they can be agents, can also be accountable moral agents' (2015: 110).

Floridi sees this as an improvement upon the Kantian position, but this means that it still rests upon some concept of the moral autonomous individual even though that is not necessarily human, and it is exactly that understanding of the human (and non-human) which New Materialism and those of us who develop this further, bring into question. The binary between agent and patient is a development of, but still within the same philosophical paradigm as a Kantian position. It is just that human characteristics are extended into the non-human, which is something that Floridi himself wants to avoid. Once again, we would argue instead that understanding humans as parts of human and non-human assemblages, and therefore also realizing that there are assemblages which don’t contain the human at all, is a more convincing means of recognizing that there is moral agency here which needs to be addressed and evaluated.

We are arguing for a modest ethics, linked to a similar modesty in both science and religion, but one has to be careful with this terminology: a modest claim is still a claim which itself suggests a degree of force in the argument. Behind this lie a series of substantive positions as articulated in the text. Knowledge as embedded and material rather than distant and abstract, taking into account non-specialist perspectives, material practices and the insights of other disciplines. A willingness to acknowledge the other levels at which humans function, those of feelings and instincts as well as what is normally termed the logical and autonomous. The realization that one is always already in relationship with the non-human in shifting and evolving assemblages. Rather than lone individuals exercising a means-ends rationality, we perceive that we operate within a distributed agency and that can mean that we are as much shaped by the technology we devise as we imagine the creation is under our complete control. Along with some of the other influences identified, such as the work of Latour and Stiegler, this can lead to a more modest ethics drawing on the insights of New Materialism as related to the themes just identified.

In terms of the specific issues involved in the development of virtual humans there is much to be debated as to how those traditional boundaries between the human and the non-human are to be reconfigured and revised. To what extent do those feelings or emotions we attribute to humans play a part both within the construction of virtual humans let alone in their relationships with humans? What may change to both through the interactions and relationships that will now develop? Does the language of rights still have a function in this new context? What exactly is sentience and how essential is it to what we will, from now on, count as authentically human? Once it is acknowledged that it is always the case that even humans operate as part of shifting and re-forming assemblages, how will this impact upon the ethical conundrums that we face as a society? How does society itself guard against the purely commercial motivations that will determine such issues?

Each of these questions requires the more fluid and dynamic interpretations of reality that can be mined from New Materialism and related ideas. Ethics is less about constructing frameworks according to which one can judge whether or not one is living a good life or taking appropriate actions, than being worthy of the events over which one has no or little control. Can we affirm, enhance and intensify such events through the capacity to affect and be affected? In order to be of some practical use, however, this requires a capacity to reflect and act, and not simply to be swept along by events. What does this look like in practice and does it have advantages over the more traditional, Kantian, rights-based approach we have also examined?

The concern with a traditional, rights-based or even environmental ethical approach is that both come too late on the scene. Such approaches apply certain ethical principles only after the important decisions have already been made: the targeting and the biases are already built into the systems being deployed. Trying to identify or even challenge and change those once they are already functioning is going to be too late. Governance procedures might limit the scope of these operations, but no more than that. The difficulty is that of entering the debate at an earlier stage, before the decisions have been made. We would argue that the understandings of a modest ethics as above stand a better chance of avoiding the instrumentalist approach of more traditional ethical frameworks—in other words, the ethics is no more than an application of ideas to systems that are already in place—and instead engage with the material energies and movements embedded in the technologies as they are being developed.

Discussion

The differences between Floridi's approach and that of a more modest ethics are that his ethics claims to be universal whereas ours is a more guarded in its claims and does not rest on a hierarchical ontology. This becomes a political as well as an ethical challenge, as illustrated in Table 1.

There is a clear difference over the issue of the material and the informational as Floridi wants to reduce everything to the informational. His stance might be better understood in the alternative framework as assemblages of the human and the non-human, and indeed simply of the non-human. The issue of autonomy versus distributed agency is also central to the argument, with the former still acting as an external and human-based understanding of the world whereas the latter offers greater roles for the non-human as agents. At the heart of the difference is Floridi's binary of agent and patient as a New Materialist approach would destabilize and disrupt such a binary and question whether this is a satisfactory means of engaging with not only other humans but also the non-human. If this could be translated into activity then it might enable that earlier intervention into the development of virtual humans and the values that will drive it that both we and Floridi would wish to see established.

What is clear from the review of the ethical conundrums is that there are two major issues. The first issue is agency. The second issue is how to engage and intervene in the developments that are yet to take place and how to influence or shape the values that will determine them. At the heart of the difference is Floridi's binary of agent and patient, as a modest ethics approach would destabilize and disrupt such a binary and question whether this is a satisfactory means of engaging with not only other humans but also the non-humans. If this could be translated into action, then it might enable that earlier intervention into the development of virtual humans and the values that will drive such development that both we and Floridi would wish to see established. Experimentation and the use of virtual humans has been discussed for many years, but the ethical concerns in this area

Table 1 Ethical conundrums

	Floridian stance	Modest ethics stance
Informational ethics	Universal	Embedded materialist
Robotic ethics	Rights based	Interactive assemblages
Affective attachment	Expanded or Inclusive Autonomy	Pre-autonomous
Ethical design	Governance and Law	Hybrid Forums
Agency	Patient model, namely Systems and Networks of Agents and their behaviour.	Distributed agency

remain troublesome. Indeed, as Riek and Howard (2014) note, in the US consumer robots developed by industry require little ethical intervention before being sold. Also there are practical concerns that still need to be addressed. For example, issues of privacy when virtual humans are being used as physically assistive robots, the loss when a virtual human is removed, the possible impact of virtual human sentience on humanity.

Conclusion

Ethics, along with both science and religion, need to exercise a degree of modesty. Claims to establish truth in some exclusive manner, and therefore at the cost of an open engagement with other contributions, need to be tempered by the recognition that no one approach has the monopoly of truth. The danger with our institutions and indeed our academic disciplines is that they tend to become so dominant and all-consuming that they begin to make inflated claims for their importance and significance. Modesty is needed in truth claims then, but also modesty in recognizing the day to day realities of our lives, the others on whom we depend, whether human or non-human, and the complexities which we all too often reduce to convenient generalities in order to manipulate and control. In philosophy there is a tendency to construct ethical frameworks in order to cope with that complexity and produce criteria by which we can evaluate human action and make decisions about the right way forward. However, one of the contributions of New Materialism is to challenge this way of working on the grounds that it oversimplifies the complexity and underestimates the uncertainties and contingencies of our lives. We need to become more comfortable with that which we cannot control and which is constantly moving beyond us. Modesty may demand that we move more slowly and carefully rather than dashing ahead in ways that our culture and digital technology encourages us to do.

References

Asimov, I. (1950). *I, robot*. New York: Doubleday.

Bennett, J. (2010). *Vibrant matter: A political ecology of things*. Durham: Duke University Press.

Borenstein, J., & Arkin, R. C. (2016). Robots, ethics, and intimacy: The need for scientific research. In D. Berkich & M. V. d'Alfonso (Eds.), *On the cognitive, ethical, and scientific dimensions of artificial intelligence: Themes from IACAP 2016* (pp. 299–309). Cham: Springer Nature. https://doi.org/10.1007/978-3-030-01800-9_16.

Braidotti, R. (2013). *The Posthuman*. Cambridge: Polity Press.

Bryant, L. (2014). *Onto-cartography: An ontology of machines and media*. Edinburgh: Edinburgh University Press.

Burden, D., & Savin-Baden, M. (2019). *Virtual humans: Today and tomorrow*. New York: Chapman and Hall/CRC.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Coole, D., & Frost, S. (Eds.). (2010). *New materialisms: Ontology, agency, and politics*. Durham, NC: Duke University Press Books.

DeLanda, M. (2002). *Intensive science and virtual philosophy*. London: Continuum.

DeLanda, M. (2010). *Deleuze: History and Science.* New York: Atropos Press.

Deleuze, G. (1987). *Dialogues*. New York: Columbia University Press.

Floridi, L. (2011). *The philosophy of information*. Oxford: Oxford University Press.

Floridi, L. (2015). *The ethics of information*. Oxford: Oxford University Press.

Harari, Y. N. (2015). *Homo Deus*. London: Harvill Secker.

Hayles, K. (1999). *How we became Posthuman: Virtual bodies in cybernetics, literature and informatics*. Chicago, IL: The University of Chicago Press.

Hayles, K. (2012). *How we think: Digital media and contemporary Technogenesis*. Chicago, IL: The University of Chicago Press.

Hegel, G. W. F. (1821). *The philosophy of right.* Translated by T. M. Knox. New York: Oxford University Press.

Hern, A. (2017). 'Give robots "personhood" status, EU committee argues. The Guardian, 12 January. https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues. Accessed 26 November 2019.

House of Lords Select Committee on Artificial Intelligence (2018). HL Paper 100 AI in the UK: ready, willing and able? House of Lords Select Committee on Artificial Intelligence Report of Session 2017–2019. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf . Accessed 26 November 2019.

Jandrić, P., Knox, J., Besley, T., Ryberg, T., Suoranta, J., & Hayes, S. (2018). Postdigital science and education. *Educational Philosophy and Theory, 50*(10), 893–899. https://doi.org/10.1080/00131857.2018.1454000.

Kant, I. (1998). *Groundwork of the metaphysics of morals.* Translated and edited by M. Gregor. Cambridge: Cambridge University Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* Milton Keynes: Open University Press.

Latour, B. (2004). *Politics of nature: How to bring the sciences into democracy.* Cambridge: Harvard University Press.

Levy, D. (2008). *Love and sex with robots: The evolution of human-robot relationships*. New York: Harper Perennial.

Malle, B. F. (2016). Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics Information Technology, 18*, 243–256. https://doi.org/10.1007/s10676-015-9367-8.

Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action: Influence of human and computer co-actors. *Experimental Brain Research, 211*(3–4), 663–670. https://doi.org/10.1007/s00221-011-2662-7.

Prinz, J. J. (2011). Is empathy necessary for morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 211–229). Oxford: Oxford University Press.

Reader, J. (2017). *Theology and new materialism: Spaces of faithful dissent*. New York: Palgrave Macmillan.

Reader, J., & Evans, A. (2019). *Ethics after New Materialism: A Modest Undertaking*. Temple ethical futures No2. Rochdale, UK: William Temple Foundation. https://williamtemplefoundation.org.uk/temple-ethical-futures/. Accessed 26 November 2019.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television and new media like real people and places*. New York: Cambridge University Press.

Riek, L.D., & Howard, D. (2014). A code of ethics for the human-robot interaction profession (April 4). Proceedings of we robot 2014. https://ssrn.com/abstract=2757805. Accessed 26 November 2019.

Savin-Baden, M., & Reader, J. (2018). Technology Transforming Theology: Digital Impacts. http://williamtemplefoundation.org.uk/our-work/temple-tracts/. Accessed 26 November 2019.

Savin-Baden, M., Tombs, G., Burden, D., & Wood, C. (2013). It's almost like talking to a person: Student disclosure to pedagogical agents in sensitive settings. *International Journal of Mobile and Blended Learning, 5*(2), 78–93. https://doi.org/10.4018/jmbl.2013040105.

Searle, J. R. (2014). Introduction: Addressing the hard problem. *Journal of Integrative Neuroscience, 13*(2), 7–11.

Stengers, I. (2011). *Thinking with whitehead*. Cambridge: Harvard University Press.

Stiegler, B. (2016). *Automatic society: The future of work.* Cambridge: Polity Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.

Winfield, A. F., Michael, K., Pitt, J., & Evers, V. (2019). Machine ethics: The design and governance of ethical AI and autonomous systems. *Proceedings of the IEEE, 107*(3), 509–517. https://doi.org/10.1109/JPROC.2019.2900622 .