Ethics and Information Technology **1:** 5–14, 1999.

The ethics of representation and action in virtual reality

Philip Brey
Twente University, Fac. WMW/SW, Enschede, the Netherlands. E-mail: p.a.e.brey@wmw.utwente.nl

Abstract. This essay addresses ethical aspects of the design and use of virtual reality (VR) systems, focusing on the behavioral options made available in such systems and the manner in which reality is represented or simulated in them. An assessment is made of the morality of 'immoral' behavior in virtual reality, and of the virtual modeling of such behavior. Thereafter, the ethical aspects of misrepresentation and biased representation in VR applications are discussed.

Introduction

Virtual reality technology, like any major new technology, raises ethical questions. So far, however, only a handful of studies have appeared that discuss ethical aspects of virtual reality.[1] As can be observed in these studies, some of the ethical questions raised by virtual reality (VR) technology concern particular uses or areas of application. For example, ethical questions can be raised concerning military applications of VR, the use of VR in therapy, or the dangers of extended uses of VR systems. VR systems may also raise ethical questions that transcend particular uses or areas of application, and that apply to VR technology in general. In this essay, I will discuss ethical questions raised by two key aspects of VR systems that transcend particular uses of them: *representational* aspects and *interactive* or *behavioral* aspects. The representational aspects of VR applications concern features of them that define the way in which objects, state-of-affairs and events are depicted or simulated. Behavioral aspects concern the actions or behaviors made possible by, or performed in, virtual reality environments.

The structure of the paper is as follows. In the next section, a definition of virtual reality technology is proposed and argued for, various kinds of virtual reality are distinguished that are relevant to the ethical analysis of VR, and various areas of application of VR systems are briefly reviewed. Section 3 contains an ethical analysis and discussion of behavior in virtual reality. Section 4 does the same for the topic of representation in VR systems. The conclusion summarizes key ethical implications, particularly those concerning the responsibility of developers of VR applications.

Virtual reality systems and areas of application

The term 'virtual reality' has no standard meaning, not even within the self-labeled virtual reality industry. There are extremely liberal uses of the term according to which any visual representational medium, including television and even paintings qualify as instances of virtual reality. More common are very restrictive uses of the term, in which virtual reality systems are defined as systems that employ head-mounted displays, datagloves and datasuits to simulate an immersive, interactive computer-generated environment. It is now recognized in the virtual reality industry that this use of the term is too restrictive, and that there are forms of virtual reality that do not involve head-mounted displays or total immersion. In *projection virtual reality*, for example, three-dimensional virtual models are projected in a room, and can be perceived from different angles by users who wear special glasses. *Desktop virtual reality* involves a virtual environment represented on a computer screen that can be perceived stereoscopically by wearing special stereoglasses. Users can interact with the represented environment with datagloves, or, more commonly, with a mouse.

In this essay, a virtual reality is defined as *a three-dimensional interactive computer-generated environment that incorporates a first-person perspective*. This means, first of all, that the attribute of full immersion is not taken to be an essential property for systems to qualify as virtual reality systems. Likewise, inter-

[1] M. Cranford. The Social Trajectory of Virtual Reality: Substantive Ethics in a World Without Constraints. *Technology in Society*, 18, 79–92, 1996; L. Whalley. Ethical Issues in the Application of Virtual Reality to Medicine. *Computers in Biology and Medicine*, 25, 107–114, 1995; S. Thurmel. Ethische Aspecte der Virtuellen Realität. In G. Meggle and U. Wessels, editors, *Analyomen 1. Proceedings of the 1st Conference on Perspectives in Analytical Philosophy*. De Gruyter, Berlin, 1994; C. Beardon. The Ethics of Virtual Reality. *Intelligent Tutoring Media*, 3, 23–28, 1992.

action through data gloves is not held to be essential, as interaction may also take place through a mouse or joystick. Stereo vision is likewise not held to be essential. Essential features of virtual reality, as defined here, are interactivity, the use of three-dimensional graphics, and a first-person perspective. *Interactivity* entails that the represented environment must allow for manipulation and navigation. Manipulation implies the modification of aspects of the environment in a fairly direct way, for example by clicking on them with a mouse or by grabbing them through the use of datagloves. Navigation implies the ability to change the location from which one perceives and manipulates features of the environment, including the ability to perceive objects from different angles. Such interactivity requires three-dimensional graphics. Text-based computer-generated environments or two-dimensional graphical environments hence do not qualify as genuine virtual realities. A *first-person perspective*, finally, entails that the environment is perceived and interacted with from a single locus. A first-person perspective suggests a degree of immersion in a world, rather than the experience of the world as an object that can be (partially) controlled from the outside.

The extent to which a virtual environment is experienced as similar to a real environment depends on the extent to which the perceptual and interactive features of the VR system or program are designed. Highly realistic VR is VR in which the user feels totally immersed, just like in a real environment, because the perceptual and interactive features of the environment have the same richness as features in a real environment. This richness is determined to the extent to which the VR system is responsive to the *sensorimotor abilities* of the human body, including perceptual abilities such as stereo vision, surround sound and tactile perception, and motor abilities such as grasping, turning one's head, and walking. In realistic VR, moreover, vision and hearing will be closed off from one's actual surroundings, usually through the use of head-mounted displays, and the user will be confined, as much as possible, to computer-generated stimuli only. In general, a realistic, highly immersive virtual environment is one that successfully simulates ordinary perceptual and motor interaction with an environment.

Currently most widespread in society are forms of VR with a relatively low degree of realism and immersion, in which environments are projected on a computer screen and controlled with a mouse, keyboard or joystick. More fully immersive forms of VR, in particular those using head-mounted displays and datagloves, are still more rare, but have in recent years gained serious professional use. For example, they are used for prototyping and ergonomic testing in industrial design, for design and design communication in architecture and urban planning, for medical and surgical simulation and training in medicine, for scientific visualization and simulation in science, for data visualization in finance, and for battlefield animation in military training. However, the use of such professional systems is still limited, and highly immersive VR also has yet to make a breakthrough as a consumer technology.

A genuine breakthrough for highly immersive VR may well occur in the near future, however, when increases in computing speed and decreases in cost bring such systems within the reach of the average consumer and make them also more appealing for business and professional use. Analysts point to the potential application areas for VR systems, which include education, training, communication (of designs and ideas), cooperative work, entertainment, and therapeutic uses, and expect VR systems to become important tools in these areas within the next ten to twenty years.[2]

In this essay, my focus will be on the ethical aspects of representation and interaction in VR systems of all varieties. However, I will not consider ethical complexities that are generated by the use of multi-user (or networked) VR systems, and confine myself to ethical aspects of single-user VR. Moreover, my greatest interest is in the ethical aspects of highly immersive VR, because its representational and interactive features tend to be more realistic than those of less immersive VR, and therefore ethical issues regarding representation and behavior in virtual reality are likely to be most pronounced in them. Some of the examples I will discuss, however, will derive from existing VR applications that are less immersive, because there do not yet exist many highly immersive VR systems that contain the morally controversial features found in these applications.

Behavior and morality in virtual reality

In virtual reality, actions may be performed that are morally proscribed in the real world. A VR application can in principle enable and graphically represent the realization of almost any imaginable action. Consequently, VR applications may be developed that allow and graphically depict almost any conceivable

[2] J. Briggs. The Promise of Virtual Reality. *Futurist*, 50, 13–18, 1996; T. Valente & T. Bardini. Virtual Diffusion or an Uncertain Reality: Networks, Policy, and Models for the Diffusion of VR Technology. In F. Biocca and M. Levy, editors, *Communication in the Age of Virtual Reality*. Lawrence Erlbaum, Hillsdale, NJ, 1995.

immoral act including murder, mutilation, torture, rape, robbery, and grand theft. Also possible in VR are role-playing and the acting out of extended scenario's, such as terrorist missions and the preparation and execution of serial killings. Currently, there are virtually no highly immersive VR applications that allow users to perform immoral actions, barring some applications in the military and medical domain. As argued in Section 2, however, it can be expected that highly immersive VR will in the future be prominent in education, training, therapy, and entertainment, and will increasingly be used to model social settings in which users interact with other virtual or real human beings. Clearly, in such applications, the possibility of unethical behavior is much more circumscribed than it is in many current applications.

With this possible future scenario in mind, I will now take a cursory look at one area in which the possibility of unethical behavior in virtual environments has already generated considerable controversy, namely in computer gaming. Many current computer games make use of realistic, three-dimensional graphics, and often employ a first-person perspective as well. They hence qualify as VR applications as defined in Section 2. Moreover, many of the highly immersive VR applications of the future may well rely in part on software and hardware technology currently used in computer gaming, as this is already an established low-cost virtual reality technology, including 3D graphics cards, joysticks, software engines for 3D modeling, inexpensive VR goggles and headsets, and network options ('multi-player support'). In what follows, I will analyze some of the actions and role-playing scenarios made possible in current 3D computer games, along with some of the social responses that such games have provoked. I will subsequently draw on this analysis for a general discussion of the ethical aspects of behavior in virtual reality environments and of the moral responsibility of their designers.[3]

- In the computer game Postal, published by Take 2 Interactive, the player personifies a crazed serial killer operating in a in a realistic neighborhood setting, evading hordes of police and killing as many innocent bystanders as possible. Actions that may be undertaken include the killing of children at school, the napalming of a marching band, and the blowing away of an anti-game-violence protest, all of which are awarded with extra points. The game also includes realistic sound fragments such as female victim's voices begging for mercy, an execute button that allows players to 'finish off' wounded people, and a suicide option, in which the killer exclaims 'I regret nothing,' before killing himself. In Australia and several other countries, the game was released in a toned-down version after censorship.
- In the computer game Grand Theft Auto, published by BMG, the player takes up the role of a small-time criminal, starting off by running errands for criminal overlords, and eventually graduating to the major league of crime, if successful at earlier stages. The game features burglary, auto theft, and driving at excessive speed to evade police, and awards extra points for actions like shooting police officers and blowing up busloads of nuns. In the United Kingdom, a ban on the game has been urged by conservative politicians.[4]
- The best-selling game Carmageddon, published by Interplay, is a hyperviolent racing game in which the goal is to smash other cars and crunch pedestrians and animals beneath one's wheels. These actions are depicted graphically and awarded with points and extra gameplay time. In the United Kingdom, a censored version was released in which the color of spilled blood was changed from red to green and human victims were made to look zombie-like. In Germany, censorship resulted in the replacement of human victims by robots. Its sequel, Carmageddon II, is more realistic than the previous version, especially in its more realistic portrayal of pedestrians and the ways in which they may be maimed and killed, and has stirred outrage even in countries in which the first version did not evoke protests.
- In the game The Wild Nines, published by Interplay, torture is one of the themes. The player acts out the role of Wex, a rebel leader in a fantasy world. The game allows the player to torture enemies for information, and depicts such torture graphically.
- The game Gender Wars, published by SCI, plays out gender differences by depicting a future society in which men and women are at war with each other. The player leads a squadron of one gender that undertakes military missions

[3] Not all these games employ a first-person perspective or allow the environment to be perceived from multiple angles. However, those that do not incorporate these features all could have been designed to incorporate them.

[4] The anti-establishment theme in the game was apparently a more important reason for urging such a ban than the violence and killing it contains. By comparison, the earlier arcade game Terminator 2, based on the movie of the same name, was opposed by police unions in the United States because it pitted players against police officers who could be shot.

against enemy soldiers or civilians of the other gender. These missions include kidnappings and assassinations of civilians, and have names like 'Castration,' and 'Die, Bastards!'.[5]

– The game Virtual Surgeon: Open Heart, published by ISM Interactive, is a so-called 'edutainment' product, in which the player learns and has fun performing an open heart operation, which is simulated in great detail. This game proves that even games that do not contain violence, sex or crime can be controversial. In the Netherlands, the game has been opposed by heart patients and their relatives, because of its game-like treatment of an emotionally charged topic.

The public debate around these hyperviolent or otherwise morally controversial computer games resembles the debate surrounding violence and morality in other media, such as television, film, and comic books. Key issues in the public discussion are whether such games should be censored, whether they are morally defensible, what harm they can inflict on their users, especially on children, and whether they will induce individuals to behave unethically in the real world. In the standard pro-censorship position, it is claimed that such games are immoral, that they hinder moral development, that they cause immoral or anti-social behavior in the real world, and that under these circumstances the state has the right to impose censorship. In the standard anti-censorship position, the libertarian viewpoint is defended that since immoral acts in a virtual environment do not cause harm to others, the decision to engage in such behavior is private, and morality of these games or the right of individuals to use them should be decided by private citizens individually and not by the state or other acting body. It is often added that there is no evidence that such games would cause individuals to act immorally in the real world, and it is sometimes claimed that such games may even be beneficial by allowing individuals to release pent-up frustrations and act out fantasies or desires that they might otherwise act out in the real world.

In the remainder of this section, my aim is to analyze some properties of behavior in virtual reality that may be relevant to this moral debate, and that may also enhance the quality of moral reflection on behavior in virtual reality in general, including VR applications in other domains. First, I want to emphasize an important difference between VR applications and other media that is insufficiently recognized in public debates. This difference is that whereas other media, such as television and literature, are passive media that are merely experienced by their users, VR applications require the user to actively engage in behavior. In VR media, the user is not a spectator but an actor. It is hence not just the experienced content of the medium that can be controversial (e.g., the experience of violence or pornography), it is also the behavioral choices offered to the user that can become an topic of moral scrutiny.

In some media, such as board games, the user is also an actor rather than just a spectator. But VR applications differ from these in that they simulate worlds that have a much great appearance of reality. There is an experienced similarity between real-world actions and actions in VR. This similarity is greatest in highly immersive VR systems. The difference between such systems and other media is that through them, one can acquire first-person experience of what it is like to perform certain immoral actions or assume criminal roles. Moreover, this experience is, as Cranford has emphasized, often unaccompanied by the threat of (real or virtual) punishment.[6] Furthermore, such systems allow one to learn the perceptual and motor skills and ways of thinking associated with such actions and roles. In short, VR media may have a separate moral status as compared to other media, because they involve agency. The moral aspects of such agency in its own right may be assessed, as well as the possible consequences of such agency for moral development and subsequent real-world behavior, due to its great similarity to real-world behavior.

So what reasons could anyone have to believe that it is wrong to murder, rape, torture or rob virtual human characters in virtual reality? I will answer this question by applying the two most influential moral approaches currently entertained in ethics, being consequentialism and Kantian duty ethics, and by drawing out under what circumstances, if any, acts in virtual reality can be labeled morally wrong by the kinds of principles proposed in these approaches.

Kantian duty ethics, first of all, upholds as the most fundamental moral principle that human beings have a duty to treat other persons with respect, that is, to treat them as ends and not as means, or to do to them as one would expect to be treated by others oneself. However, a virtual person is not by any measure a real person, but is merely a simulation of a person, so it would seem that human beings have no intrinsic duty to treat virtual persons with respect. Yet, perhaps it can be argued that our duty to treat real persons with respect

[5] In a banned advertisement for the game, a male character was depicted on top of a pile of female bodies with the caption 'How many women did you kill today?'

[6] M. Cranford. The Social Trajectory of Virtual Reality: Substantive Ethics in a World Without Constraints. *Technology in Society*, 18, 79–92, 1996.

requires that we do not treat virtual beings or things disrespectfully. But what argument could support such a claim? As far as I can see, two arguments can be adduced in favor of it, which I will call the *argument from moral development* and the *argument from psychological harm*.

The first of these arguments can be acquired from Kant himself. Kant, along with other thinkers like Augustine and Descartes, claimed that although animals have no intrinsic worth and therefore do not deserve our respect, we should avoid treating them cruelly, because this may lead us to treat our fellow human beings cruelly as well. The reason for this is that the emotions appealed to in the treatment of animals are the same emotions that are appealed to in the treatment of fellow human beings, because these actions resemble each other so closely. Cruel treatment of animals will therefore make us less kind and more harsh in our dealings with fellow human beings. As Kant put it, "[...] he who is cruel to animals becomes hard also in his dealing with men," and "Tender feelings towards dumb animals develop humane feelings towards mankind."[7] Certainly, if disrespectful treatment of animals causes disrespectful treatment of human beings, then disrespectful treatment of virtual characters, which may be even more similar to such treatment of real humans, will have the same consequence. It should also be clear, however, that this arguments is in need of sound empirical support. Empirical evidence is needed that humans are psychologically structured such that cruel or otherwise immoral behavior practiced in one domain necessarily carries over to other, similar domains. As long as this empirical case is not made, the argument is inconclusive.

The second argument, the argument from psychological harm, is that third parties may suffer psychological harm by the knowledge that a representation of themselves or individuals like them, or representations of other beings or things that they value, are not treated with respect by others. According to this argument, people tend to identify with representations of themselves or of social categories in which they fit or with which they identify. If such representations are not treated with respect, then they themselves feel disrespected or abused. So for example, heart patients may identify with the generic heart patient represented in Virtual Surgeon, and may feel that the game-like character of the application is inconsiderate to them: The virtual operation that can be performed is not intended to teach medical students to save lives, but for the general public to have a good time. In other words, in the game heart patients function as a means, and not as an end.

It may be argued in return, however, that perhaps people should learn not to identify with such representations, and should recognize that any action performed on representations is irrelevant to events in the real world. However, this touches on a major point of controversy in Western liberal societies. In most late-twentieth century Western societies, the moral principle that has become dominant is the principle that individual actions should be allowed as long as they do not harm others. What should be the moral status of actions, however, that do not harm someone's physical integrity, personal property or social status, but that harm someone by offending his or her sensibilities? This requires a careful trade-off between the right to act and the right not to be offended. This trade-off cannot be made *a priori*, but requires a careful consideration of the arguments of the actor for his freedom to act, and of the offended party of why the behavior is offensive.

Mutual understanding of each other's motives and beliefs is moreover crucial. Specifically, whether immoral behavior in virtual reality may become acceptable to the offended party may well depend on his or her assessment of the intentions, values and beliefs of the actor. What may have to be reestablished for the offended party is a basic trust that the desire to act immorally in virtual environments does not reflect a fundamental disrespect for the real-life equivalents of the virtual beings or things that are harmed or desecrated in VR.

These two last arguments, the argument from moral development and the argument from psychological harm, are, with some adaptations, also the two arguments that are most likely to be used to support a consequentialist argument against immoral behavior in VR. Consequentialist theories of morality typically hold that those actions should be performed that bring about the greatest good over bad for everyone affected by the act, and that immoral actions are those that unnecessarily harm others. The argument from moral development, reconstructed in a consequentialist framework, may be adapted to state that immoral behavior in VR leads to actions in the real world that have harmful consequences (as compared to actions that are disrespectful to others). Just like its Kantian counterpart, this argument is suggestive, but awaits further empirical evidence. The argument from psychological harm, in its consequentialist version, is that actions in VR are immoral if the psychological harm experienced by those who are offended by such actions is greater than the joy experienced by those performing them. As with its Kantian counterpart, I want to suggest that this is a matter ultimately to be

[7] I. Kant. Duties to Animals and Spirits. In *Lectures on Ethics* (trans. L. Infield). Methuen & Company, London, 1963.

resolved in social negotiation and dialogue, not in *a priori* ethics.

Finally, I want to relate the above discussion of moral behavior to the *design* of VR systems. How may the design of a VR application determine or influence the actions that users perform in its virtual environment and the way these users experience these actions and learn from them? Designers of VR applications have several means at their disposal to exert such influence. First, of course, they determine what actions can be performed at all. Whether killing or theft is possible in a VR application depends on whether it has been programmed to be possible. Second, designers determine how actions and their consequences are represented, and which consequences are represented at all. They determine, for example, whether killings are represented in graphic detail, or what ecological consequences are displayed when the user pollutes a river. Third, designers may stimulate or induce behavior by explicitly suggesting it through text or symbols displayed or narrated in the simulation or by highlighting objects in it, or by rewarding such actions with points or other tokens of social approval, and actions may be discouraged in similar ways.

I hold that designers of VR systems have the moral duty to reflect on the way behavioral options and their consequences are designed in them, because of the morally controversial nature of this topic. Specifically, designers should ask themselves two questions regarding these behavioral options in design. First, what kinds of patently immoral actions are made possible within the generated virtual environment of the application, and should such actions made to be possible at all? Second, how is the application structured so as to encourage or prevent, or approve or disapprove of, such actions, and are alternative courses of action made possible? Designers may want to avoid encouraging users to engage in unethical behavior, children and adolescents in particular, but may also want to avoid paternalism by disallowing such behavior, especially for adult users. After all, if users of VR applications have no choice but to behave morally, then they are not free agents, and it is by now an accepted truth in moral philosophy that moral acting presupposes free agency. They may want to take care to portray the consequences of actions realistically, so as to properly inform users of the consequences of their actions when performed in the real world. And in some applications (e.g., applications for use by minors) they may want to make use of various means to communicate approval or disapproval for actions undertaken.

Currently, moral debate concerning the design of interactive features in VR applications is largely confined to computer games, but as I have argued in Section 2, important future application areas of VR may also include education, training, and therapy. Applications in these areas will be explicitly designed to make people learn things and to modify their beliefs and behavior. In such applications, the design of their interactive features (what actions are made possible, how consequences of actions are represented, and what sorts of approval and disapproval of actions are signaled to users) is even more important than in entertainment applications. Such applications, after all, are supposed to equip people with skills and knowledge that will most likely be applied in real life. So for example, a combat simulation program that requires users to kill opponents rather than shoot them in the legs or arms to disable them teaches a particular method of disabling opponents without leaving users to freedom to explore alternative methods. In general, if it is true that people learn from trying out different behaviors and observing their consequences, then the determination of behavioral options made available in VR applications and the representation of consequences of actions constitute issues that developers of VR applications should think about very seriously.

Misrepresentation and biased representation in virtual reality systems

In the discussion of the ethics of design of VR applications near the end of the previous section, the focus was on the programming of the interactive features of a virtual environment. In this section, the emphasis is on its representational features. The two are, of course, interrelated, as represented items in VR environments are (usually) interactive, and interaction with a VR world requires representation. Nevertheless, the emphasis in this section is on representation alone. The question I will be answering is whether the way in which objects, events, persons and places are represented in VR may raise any ethical questions. I will answer this question by focusing on the standards of accuracy or realism by which virtual environments are modeled and on the possibility that they contain individual or social biases.

The relation between virtual reality worlds and the real world may range from faithful correspondence to no correspondence. At one extreme, there are VR models that have been designed to faithfully simulate existing structures, state-of-affairs, or events. For example, VR applications have been developed that simulate in great detail existing buildings such as the Louvre or Taj Mahal, the behavior of existing industrial complexes or the medical condition of particular patients. At another extreme, there are

VR worlds such as those found in games that play in fantasy worlds, that have no intended correspondence with any objects or properties in the real world, and may even break with existing physical laws. In between these extremes, one finds VR applications with varying degrees of realism. Many applications aim to model generic types of structures or events without necessarily intending any reference to a particular existing structure or event. For example, a VR simulation of military combat may contain realistic portrayals of people, weaponry and natural landscapes without intending to represent particular individuals or a particular landscape. Popular simulation software like SimCity and SimLife, which simulate city development and biological processes, are supposed to be realistic but do not aim to model any particular existing cities or biological organisms.

VR applications hence differ in the kinds of *reality claims* they make, i.e., the implicit or explicit promises about the realism of (features of) the virtual environment. When certain reality claims are made, the application can be expected to live up to certain *standards of accuracy*. Standards of accuracy are standards by which it is determined what features found in an actual phenomenon should minimally be represented in the simulation, what amount of detail in the representation of these features is minimally acceptable, and what kinds of idealizations may be maintained in the simulation. In practice, standards of accuracy are defined by the purpose or function that is set for the VR application together with any further promises or claims about the level of realism of the application. So a VR application may fail to be accurate either because it is insufficiently accurate for the purpose it has been designed to serve, or because it fails to adhere to additional standards of accuracy that it is claimed to adhere to.[8]

[8] It is not just the case that standards of accuracy are weaker or stronger for different applications. They can also be *different*. for different applications. In other words, there are not just variations in the *degree* of accuracy required for a VR application, but also in the *kind* of accuracy required. This is so because different purposes served by applications may result in standards of accuracy that emphasize different dimensions or aspects of the phenomenon that is simulated. For example, a VR simulation of the Louvre implies different standards of accuracy depending on the use to which the simulation is put. If the application is used by students of architecture, then high standards of accuracy are required for the representation of the architectural features of the Louvre, whereas other features may be represented inaccurately or simply be left out. If the application is supposed to provide a virtual tour of the Louvre, then its interior decoration should be rendered with great accuracy. If the application is supposed to aid in a review of safety features in the Louvre, then these should be rendered with great accuracy. And so on.

When a VR application fails to uphold accepted standards of accuracy by representing features as real that by such standards cannot justifiably be held to be present in reality or by failing to represent features that ought to be present in the application, we may say that the application *misrepresents* reality. For example, if great accuracy in the representation of shape and color belongs to the standards accepted for a VR application, and the application represents shapes and colors different from those found in the modeled reality, then the application misrepresents reality. Verifiable misrepresentation requires that there are unambiguous, shared standards of accuracy in place according to which judgments of misrepresentation can be made.

Such shared standards of accuracy are not always present, however. There may be diverging opinions regarding the purposes or functions that should be served by a VR application, as well as disagreements regarding the standards of accuracy appropriate for applications that serve these functions. Many modeling choices in the design of virtual environments are underdetermined by empirical reality and rest in part on *pragmatic criteria* regarding the *practical use or value* of available modeling options. These pragmatic criteria are necessarily derived from the values and interests of the developers themselves, as well as the values and interests of those individuals and groups that they feel obligated to represent. Representational aspects of VR simulations that depend on pragmatic choices include at least the following:

- *The inclusion or exclusion of real-world objects or features in the simulation.* A VR simulation of a building, for example, may selectively leave out graffiti or cracks in walls, while representing frescos on it in great detail. A VR simulation of bombing raid may represent military structures in an area but fail to represent nearby civilian structures. A VR simulation of blood vessels in the brain to serve as an aid in brain surgery may leave out smaller vessels.
- *Choices concerning the level of detail and realism by which objects or features included in the simulation are represented.* For example, people and animals in VR simulations may be programmed to look realistic to different degrees and the repertoire of behaviors they display may be realistic to different degrees as well. In a VR representation of combat, wounds may be represented as faint dots, or may be represented in graphic detail.
- *The selective misrepresentation or idealized representation of objects or features for pragmatic purposes, including their representation by means of icons or symbols.* For example, for the

communication of design, an idealized model of an industrial complex that simplifies its structure may be more useful than a model that faithfully reproduces every curve and connector piece. In a simulation of surgery, organs and tissues may be rendered in unrealistic colors to facilitate recognition. In a simulation of combat, it may be convenient to remove targets that have been killed or destroyed from the simulation because they no longer serve a purpose in it. Generally, also, objects, texts and symbols in VR environments that are important for the purpose of the application may be foregrounded or highlighted so as to render them more visible to the user.

- *The use of stereotypes in the representation of people, things, and events.* Stereotypical representations of generic situations may not be factually incorrect, in that they might exist in reality as they are represented, but may rest on stereotypes that are not representative of reality at large. In particular, the representation of individuals and their behaviors and cultural artifacts may be based on ethnic or gender stereotypes. The representation of other phenomena like animals and natural environments may draw on stereotypes as well.
- *Built-in assumptions about cause-and-effect relationships and implicit narratives.* VR simulations include principles laws that the modeled environment or objects in it obey to, including a number of causal laws that define cause-and-effect relationships. Some of these laws may be difficult to verify and their inclusion is therefore dependent on personal conviction. For example, a simulation program like SimCity, which simulates city development, contains all kinds of assumptions about causal relationships, for example assumptions about the relation between poverty and crime, that are controversial and difficult to prove. Relatedly, VR simulations may by their sequencing of events contain implicit narratives or story lines. For example, a combat simulation may be structured such that any disobedience of orders ultimately results in the shooting of captivation of its participants.
- *Choices in the visualization of abstract domains*. Data visualization and the simulation of abstract domains require even more pragmatic choices and conventions than the simulation of concrete, observable domains. For example, it involves choices on what data to use, what categories to group data in, what basic modeling parameters to use, and what symbolisms (e.g., color, light intensity) to employ in the representation of abstract features.

Any VR simulation necessarily entails some or all of these pragmatic modeling choices. Consequently, any VR simulation is to a significant extent value-laden. This value-ladenness does not in itself make VR applications morally controversial. Value choices in VR simulations are only morally controversial, I claim, when these choices disadvantage certain individuals and groups that can justifiably be claimed to have a stake in the simulation because they hold certain values or have certain interests that may be compromised or may fail to be promoted by the way the simulation is set up. Particularly problematical are VR applications used for learning or training purposes that promote stereotypes, obscure consequences of actions, and suggest causal relations that are in fact politically controversial. When a VR application favors certain values or interests over others due to its choices in representation, it may be said that the model makes use of *biased representations*.[9]

Hence I distinguish two types of representational failures or shortcomings in VR simulations: *misrepresentation*, in which some aspects of the simulation are clearly mistaken according to accepted standards of accuracy, and *biased representation*, in which the values or interests of some stakeholders of the simulation are insufficiently accounted for. The principal moral importance of these representational failures is that they may induce false or biased beliefs in users that may ultimately have undesirable practical consequences. A secondary issue is that representational biases may make VR applications less useful or agreeable to users whose values or interests are disregarded in the application.

What moral responsibilities do VR modelers have regarding the possibility of misrepresentation and biased representation in their simulations? In a discussion of mathematical models, which may also be extended to VR models, Richard Mason claims: 'At a minimum, a model builder is obliged to do at least three things: (1) to represent reality to clients adequately, (2) to understand and to incorporate the clients' values into the model in an effective way, and (3) to ensure that actions the client takes based on the model have the desired effect.'[10] I agree with this assessment, with the addition that from a moral point of view, it is not just the clients' values and desires

[9] It may be noted that representational biases are not the only biases that can be found in VR applications. There may also be *interactive biases*, biases in the interactive possibilities offered by the simulation, that make the simulation less useful to, or less compatible with the values of, some individuals or groups, or that selectively encourage, reward, disallow, or punish, certain actions or behaviors in virtual reality. See the previous section.

[10] M. Mason. Morality and Models. In W. Wallace, editor, *Ethics in Modeling*. Elsevier Science, Oxford, 1994.

that are relevant in the assessment of a model, but the values and interests of other stakeholder of the model as well. By dropping Mason's third principle, which I think is implicit in his first two principles, his principles can then be formulated as follows: A model builder is obligated to (1) avoid misrepresentation in his models, and (2) to avoid biased representation in his models, by recognizing stakeholders of the model and by taking into account their values and interests.

The responsibility to avoid misrepresentation entails the responsibility to take proper precautions to ensure that modeling mistakes do not occur, especially when the stakes are high (e.g., in professional simulations of surgery, or design models of collapsible structures like bridges and buildings). It also entails the responsibility to inform users if such mistakes do occur and are difficult to correct. And, finally, it entails the responsibility not to participate in intentional deception of users (e.g., embellishment, dramatization, virtual censorship). These responsibility can be derived from general responsibilities of engineers, as found in professional codes, to hold paramount the safety, health and welfare of the public, to inform clients and other stakeholders of engineering mistakes, and to act in an objective and truthful way.

The responsibility to avoid biased representation can be derived from the general responsibility of engineers (and other professionals) to use their knowledge and skill for the enhancement of human welfare. Certainly, it is in the interest of human welfare that VR applications do not contain biases that disregard the values and interests of some of their stakeholders. Developers of VR applications should therefore take care to recognize how biases may enter into their designs and take steps to avoid such biases. This will require the development of a methodology for the recognition and avoidance of such biases. Although a developed methodology does not currently exist, Friedman and Nissenbaum is a good starting point, as it analyzes various kinds of biases that may emerge at different stages in the design of computer systems.[11]

It may be noted, finally, that many of the above remarks on misrepresentation and biased representation also apply to other representational media. Paintings can be inaccurate, graphs can be misleading, documentaries may be biased, etc. An important difference between such media and VR, however, is that the representations of VR media present themselves as elements in full-blown, three-dimensional, interactive environments. VR simulations of objects may approach the perceptual complexity and interactive richness of everyday physical objects, and may for this reason more easily generate belief in their veracity and objectivity than other sorts of representations. This effect may be strongest for VR simulations of abstract domains. Financial data, for example, may be harder to disbelieve when one can hold it in one's hand or navigate through it than when it is displayed as numbers on a sheet of paper. However, whether VR media indeed have this effect of more easily inducing belief in their veracity than other media requires further empirical study.

[11] B. Friedman and H. Nissenbaum. Bias in Computer Systems. In B. Friedman, editor, *Human Values and the Design of Computer Technology*. Cambridge University Press, Cambridge, 1997.

Conclusion

The representational and interactive features of VR applications have been argued to raise significant ethical questions. VR applications, especially those aimed at realistic simulation, may involve misrepresentation and biased representation, and developers have been argued to have the responsibility to take proper precautions to avoid misrepresentation or inform users about its occurrence, and to chart the stakeholders of their applications and ensure that their values and interests are accounted for in design choices regarding representational format and content. Behavior in virtual environments also raises ethical questions. These questions have, however, proven to be more difficult to resolve. A more extended discussion is needed of the ethical aspects of immoral behavior in VR environments, along with more empirical research on the implications for real-world behavior of behavior in virtual worlds. Designers were claimed to have the responsibility to reflect on their own standpoint on this issue, and to reflect on the moral aspects of the way in which behavioral options and the consequences of actions are structured and represented in VR applications. This has been argued to be especially important in VR systems used in education, training and therapy, and applications for use by children and adolescents.

References

C. Beardon. The Ethics of Virtual Reality. *Intelligent Tutoring Media* 3: 23–28, 1992.

J. Briggs. The Promise of Virtual Reality. *Futurist* 50: 13–18, 1996.

M. Cranford. The Social Trajectory of Virtual Reality: Substantive Ethics in a World Without Constraints. *Technology in Society* 18: 79–92, 1996.

B. Friedman and H. Nissenbaum. Bias in Computer Systems. In B. Friedman, editor, *Human Values and the Design of Computer Technology*. Cambridge University Press, Cambridge, 1997.

I. Kant. Duties to Animals and Spirits. In *Lectures on Ethics* (trans. L. Infield). Methuen & Company, London, 1963.

M. Mason. Morality and Models. In W. Wallace, editor, *Ethics in Modeling*. Elsevier Science, Oxford, 1994.

S. Thurmel. Ethische Aspecte der Virtuellen Realität. In G. Meggle and U. Wessels, editors, *Analyomen 1. Proceedings of the 1st Conference on Perspectives in Analytical Philosophy*. De Gruyter, Berlin, 1994

T. Valente and T. Bardini. Virtual Diffusion or an Uncertain Reality: Networks, Policy, and Models for the Diffusion of VR Technology. In F. Biocca and M. Levy, editors, *Communication in the Age of Virtual Reality*. Lawrence Erlbaum, Hillsdale, NJ, 1995.

L. Whalley. Ethical Issues in the Application of Virtual Reality to Medicine. *Computers in Biology and Medicine* 25: 107–114, 1995.