



2-19-2016

Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology

Michael Madary

University of the Pacific, mmadary@pacific.edu

Thomas K. Metzinger

Johannes Gutenberg – Universität Mainz

Follow this and additional works at: <https://scholarlycommons.pacific.edu/cop-facarticles>



Part of the [Philosophy Commons](#)

Recommended Citation

Madary, M., & Metzinger, T. K. (2016). Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI*, 3(3), DOI: [10.3389/frobt.2016.00003](https://doi.org/10.3389/frobt.2016.00003)
<https://scholarlycommons.pacific.edu/cop-facarticles/754>

This Article is brought to you for free and open access by the All Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in College of the Pacific Faculty Articles by an authorized administrator of Scholarly Commons. For more information, please contact mgibney@pacific.edu.



Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology

Michael Madary* and Thomas K. Metzinger

Johannes Gutenberg – Universität Mainz, Mainz, Germany

OPEN ACCESS

Edited by:

Nadia Magnenat Thalmann,
University of Geneva, Switzerland

Reviewed by:

Jean-Marie Normand,
Ecole Centrale de Nantes, France
HyungSeok Kim,
Konkuk University, South Korea
Roland Blach,
Fraunhofer Institut für
Arbeitswirtschaft und Organisation,
Germany

*Correspondence:

Michael Madary
madary@uni-mainz.de

Specialty section:

This article was submitted to
Virtual Environments,
a section of the journal
Frontiers in Robotics and AI

Received: 05 December 2015

Accepted: 08 February 2016

Published: 19 February 2016

Citation:

Madary M and Metzinger TK (2016)
Real Virtuality: A Code of Ethical
Conduct. Recommendations for
Good Scientific Practice and the
Consumers of VR-Technology.
Front. Robot. AI 3:3.
doi: 10.3389/frobt.2016.00003

The goal of this article is to present a first list of ethical concerns that may arise from research and personal use of virtual reality (VR) and related technology, and to offer concrete recommendations for minimizing those risks. Many of the recommendations call for focused research initiatives. In the first part of the article, we discuss the relevant evidence from psychology that motivates our concerns. In Section “Plasticity in the Human Mind,” we cover some of the main results suggesting that one’s environment can influence one’s psychological states, as well as recent work on inducing illusions of embodiment. Then, in Section “Illusions of Embodiment and Their Lasting Effect,” we go on to discuss recent evidence indicating that immersion in VR can have psychological effects that last after leaving the virtual environment. In the second part of the article, we turn to the risks and recommendations. We begin, in Section “The Research Ethics of VR,” with the research ethics of VR, covering six main topics: the limits of experimental environments, informed consent, clinical risks, dual-use, online research, and a general point about the limitations of a code of conduct for research. Then, in Section “Risks for Individuals and Society,” we turn to the risks of VR for the general public, covering four main topics: long-term immersion, neglect of the social and physical environment, risky content, and privacy. We offer concrete recommendations for each of these 10 topics, summarized in **Table 1**.

Keywords: ethics, virtual reality, augmented reality, substitutional reality, depersonalization disorder, derealization, informed consent, dual use

PRELIMINARY REMARKS

Media reports indicate that virtual reality (VR) headsets will be commercially available in early 2016, or shortly thereafter, with offerings from, for example, Facebook (Oculus), HTC and Valve (Vive) Microsoft (HoloLens), and Sony (Morpheus). There has been a good bit of attention devoted to the exciting possibilities that this new technology and the research behind it have to offer, but there has been less attention devoted to novel ethical issues or the risks and dangers that are foreseeable with the widespread use of VR. Here, we wish to list some of the ethical issues, present a first, non-exhaustive list of those risks, and offer concrete recommendations for minimizing them. Of course, all this takes place in a wider sociocultural context: VR is a technology, and technologies change the objective world. Objective changes are subjectively perceived, and may lead to correlated shifts in value judgments. VR technology will eventually change not only our general image of humanity but also our

understanding of deeply entrenched notions, such as “conscious experience,” “selfhood,” “authenticity,” or “realness.” In addition, it will transform the structure of our life-world, bringing about entirely novel forms of everyday social interactions and changing the very relationship we have to our own minds. In short, there will be a complex and dynamic interaction between “normality” (in the descriptive sense) and “normalization” (in the normative sense), and it is hard to predict where the overall process will lead us (Metzinger and Hildt, 2011).

Before beginning, we should quickly situate this article within the larger field of the philosophy of technology. Brey (2010) has offered a helpful taxonomy dividing the philosophy of technology into the classical works from the mid-twentieth century, on the one hand, and more recent developments that follow an “empirical turn” by focusing on the nature of particular emerging technologies, on the other hand. We intend the present article to be a contribution to the latter kind of philosophy of technology. In particular, we are investigating foundational issues in *the applied ethics of VR*, with a heavy emphasis on recent empirical results. Both authors have been participants in the collaborative project *Virtual Embodiment and Robotic Re-Embodiment* (VERE), a 5-year research program funded by the European Commission.¹ Despite this explicit focus, we do not mean to imply that the issues investigated here will not find fruitful application to themes from classical twentieth century philosophy of technology (see Franssen et al., 2009). Consider, for instance, Martin Heidegger’s influential treatment of the way in which modern technology distorts our metaphysics of the natural world (Heidegger, 1977; also Borgmann, 1984), or Herbert Marcuse’s prescient account of industrial society’s ongoing creation of false needs that undermine our capacities for individuality (Marcuse, 1964). As should become clear from the examples below, immersive VR introduces new and dramatic ways of disrupting our relationship to the natural world (see Neglect of Others and the Physical Environment). Likewise, the newly created “need” to interact using social media will become even more psychologically ingrained as the interactions begin to take place while we are embodied in virtual spaces (see The Effects of Long-Term Immersion and O’Brocháin et al., 2016). In sum, the fact that connections with classical philosophy of technology will remain largely implicit in this article should not be taken to suggest that they are not of great importance.

The main focus will be on immersive VR, in which subjects use a head-mounted display (HMD) to create the feeling of being within a virtual environment. Although our main topic involves the experience of immersion, some of the concerns raised, such as neglect of the physical environment (see Neglect of Others and the Physical Environment), can be applied to extended use of an HMD even when users do not experience immersion such as

when merely using the device for 3D viewing. Many of our points are also relevant for other types of VR hardware, such as CAVE projection. One central area of concern has to do with *illusions of embodiment*, in which one has the feeling of being embodied other than in one’s actual physical body (Petkova and Ehrsson, 2008; Slater et al., 2010). In VR, for instance, one might have the illusion of being embodied in an avatar that looks just like one’s physical body. Or one might have the illusion of being embodied in an avatar of a different size, age, or skin color. In all of these cases, insight into the illusory nature of the overall state is preserved. The fact that VR technology can induce illusions of embodiment is one of the main motivations behind our investigation into the new risks generated by using VR by researchers and by the general public. Traditional paradigms in experimental psychology cannot induce these strong illusions. Similarly, watching a film or playing a non-immersive video game cannot create the strong illusion of owning and controlling a body that is not your own. Although our main focus will be on VR (see **Figure 1**), many of the risks and recommendations can be extended to augmented reality (Azuma, 1997; Metz, 2012; Huang et al., 2013 and substitutional reality (Suzuki et al., 2012; Fan et al., 2013). In augmented reality (AR, see **Figure 2**), one experiences virtual elements intermixed with one’s actual physical environment.

Following Milgram and colleagues (Milgram and Kishino, 1994; Milgram and Colquhoun, 1999), it may be helpful here to consider augmented reality along the Reality–Virtuality Continuum. The real environment is located at one extreme of the continuum and an entirely virtual environment is located at the other extreme. Displays can be placed along the continuum according to whether they primarily represent the real environment while including some virtual elements (augmented reality) or they primarily represent a virtual environment while including some real elements (augmented virtuality). Much of the following discussion will focus on entirely virtual environments, but readers should keep in mind that many of the concerns raised will also apply to environments all along the Reality–Virtuality Continuum.

It is foreseeable that there will be ever new extensions and special cases of VR. We return to this theme with some philosophical remarks at the end of the article. For now, let us at least note that the very distinction between the real and the virtual is ripe for further philosophical investigation. One example of such a special, recent extension of VR that does not in itself form a distinct new category is “substitutional reality” (SR, see **Figure 3**), in which an omni-directional video feed gives one the illusion of being in a different location in space and/or time, and insight may not be preserved. Readers should keep in mind that VR headsets will likely enable users to toggle between virtual, augmented, and substitutional reality, and to adjust one’s location on the Reality–Virtuality Continuum, thus somewhat blurring the boundaries between kinds of immersive environments.

We divide our discussion into two main areas. First, we will address the research ethics of VR. Then we will turn to issues arising with the use of VR by the general public for entertainment and other purposes. To be clear upfront, we are not calling for general restrictions on an individual’s liberty to spend time (and money) in VR. In open democratic societies, such regulations must be based on rational arguments and available empirical evidence,

¹The project, as well as the current publication, is funded under the EU 7th Framework Program, Future and Emerging Technologies (Grant 257695). VERE aimed at dissolving the boundary between the human body and surrogate representations in immersive virtual reality and physical reality, giving people the illusion that their surrogate representation is their own body. See <http://www.vereproject.eu/> for more. We thank members of the VERE consortium for discussing many of the issues in this article during our VERE Ethics Workshops in February 2013 and September 2015.



FIGURE 1 | Illusory ownership of an avatar in virtual reality. Here, a subject is shown wearing a head-mounted display and a body tracking suit. The subject can see his avatar in VR moving in synchrony with his own movements in a virtual mirror. In this case, the avatar is designed to replicate Sigmund Freud in order to enable subjects to counsel themselves. Thus, creating what Freud may have called an instance of avatar-introjection! (Image used with kind permission from Osimo et al., 2015.)

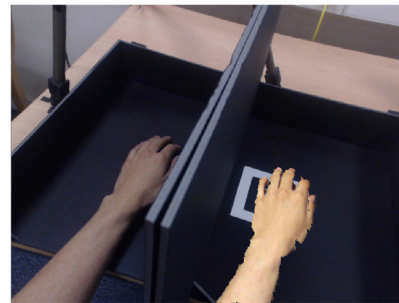
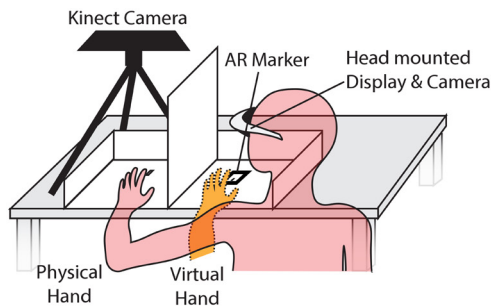


FIGURE 2 | An augmented reality hand illusion. Here, augmented reality is used to show the subject a virtual hand in a biologically realistic location relative to his own body. This case differs from virtual reality due to the fact that the subject sees the virtual hand embedded in his own physical environment rather than in an entirely virtual environment (image used with kind permission from Keisuke Suzuki).

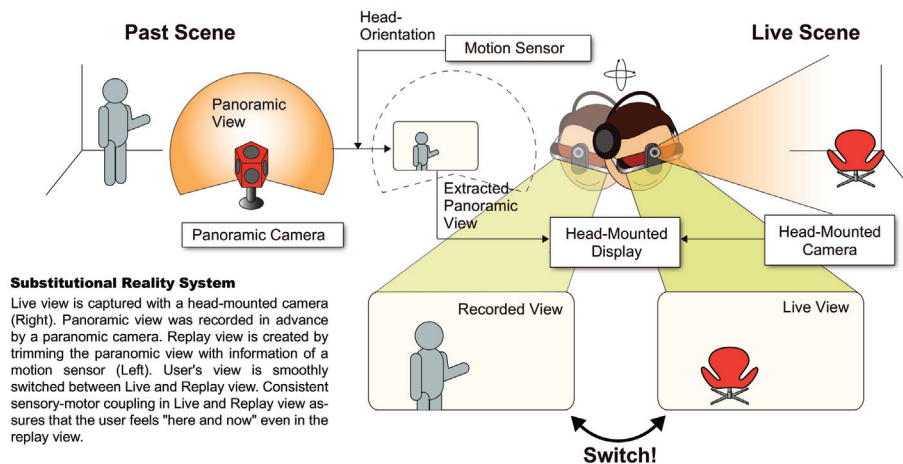


FIGURE 3 | Immersion in the past using substitutional reality. In this example, substitutional reality is used to allow switching between a live view of the scene and a panoramic recording of that scene from the past. Note that SR could also be used to provide live (or recorded) panoramic input from a distant location, creating the illusion that one is "present" somewhere else (image used with kind permission from Anil Seth).

and they should be guided by a general principle of liberalism: in principle, the individual citizen's freedom and autonomy in dealing with their own brain and in choosing their own desired states of mind (including all of their phenomenal and cognitive properties) should be maximized. As a matter of fact, we would even argue for a constitutional right to mental self-determination (Bublitz and Merkel, 2014), somewhat limiting the authority of the government, because the above-mentioned values of individual freedom and mental autonomy seem to be absolutely fundamental to the idea of a liberal democracy involving a separation of powers. However, once such a general principle has been clearly stated, the much more interesting and demanding task lies in helping individuals exercise this freedom in an intelligent way, in order to minimize potential adverse effects and the overall psychosocial cost to society as a whole (Metzinger, 2009a; Metzinger and Hildt, 2011). New technologies like VR open a vast space of potential actions. This space has to be constrained in a rational and evidence-based manner.

Similarly, we fully support ongoing research using VR – indeed, we argue below that there are ethical demands to do *more* research using it, research that is motivated in part with the goal of mitigating harm for the general public. But we do think that it is prudent to anticipate risks and we wish to spread awareness of how possibly to avoid, or at least minimize, those risks.² Before entering into the concrete details, we are going to make the case for being especially concerned about VR technology in contrast, say, to television or non-immersive video games. We do so in two steps. First, in Section “Plasticity in the Human Mind,” we cover some of the relevant discoveries from psychology in the past decades, including the scientific foundation for illusions of embodiment. Then in Section “Illusions of Embodiment and Their Lasting Effect,” we cover the more recent experimental work that has begun to reveal the lasting psychological effects of these illusions. Then in Section “Recommendations for the Use of VR by Researchers and Consumers,” we will cover the research ethics of VR followed by risks for the general public.

Plasticity in the Human Mind

One central result of modern experimental psychology is that human behavior can be strongly influenced by external factors while the agent is totally unaware of this influence. Behavior is context sensitive and the mind is plastic, which is to say that it is capable of being continuously shaped and re-shaped by a host of causal factors. These results, some of which we present below, suggest that our environment, including technology and other humans, has an unconscious influence on our behavior. Note that the results do not conflict with the manifest fact that most of us have relatively stable character traits over time. After all, most of us spend our time in relatively stable environments. And there may be many aspects of the functional architecture

²Behr et al. (2005) have addressed similar themes about practical issues in VR research and applications. Here, we wish to address concerns that go beyond their initial treatment of the topic. More recently, O’Brocháin et al. (2016) have covered ways in which the conjunction of VR with social networks might raise threats to privacy and autonomy. We will engage with some of their concerns at various points below.

underlying the neurally realized part of the human self-model [for example, of the body model in our brain, e.g., Metzinger (2003), p. 355] that are largely genetically determined. However, we also want to point out that human beings possess a large number of *epigenetic* traits, that is, a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence.

Context-Sensitivity All the Way Down

The way in which our behavior is sensitive to environmental features is especially relevant here due to the fact that VR introduces a completely new type of environment, a new cognitive and cultural niche, which we are now constructing for ourselves as a species.

It is not excluded that extended interactions with VR environments may lead to more fundamental changes, not only on a psychological, but also on a biological level.

Some of the most famous experiments in psychology reveal the context sensitivity of human behavior. These include the Stanford Prison Experiment, in which normal subjects playing roles as either prison guards or inmates began to show pathological behavioral traits (Haney et al., 1973), Milgram’s obedience experiments, in which subjects obeyed orders that they believed to cause serious pain and be immoral (Milgram, 1974), and Asch’s conformity experiments, in which subjects gave obviously incorrect answers to questions after hearing confederates, all give the same incorrect answers (Asch, 1951). For a more recent result showing the unconscious impact of environment on behavior, the amount of money placed in a collection box for drinks in a university break room was measured under a condition in which the image of a pair of eyes was posted above the collection box. With the eyes “watching,” coffee drinkers placed three times as much money in the box compared to the control condition with no eyes (Bateson et al., 2006). Effects like this one may be particularly relevant in VR, because the subjective experience of presence and *being* there is not only determined by functional factors like the number and fidelity of sensory input and output channels, the ability to modify the virtual environment, but also, importantly, the level of social interactivity, for example, in terms of actually being recognized as an existing person by *others* in the virtual world (Heeter, 1992; Metzinger, 2003). As investigations into VR have interestingly shown, a phenomenal reality as such becomes more *real* – in terms of the subjective experience of presence – as more agents recognizing one and interacting with one are contained in this reality. Phenomenologically, ongoing social cognition enhances both this reality and the self in their degree of “realness.” This principle will also hold if the subjective experience of ongoing social cognition is of a hallucinatory nature.

Potential for Deep Behavioral Manipulation

Whether physical or virtual, human behavior is situated and socially contextualized, and we are often unaware of the causal impact this fact has on learning mechanisms as well as on

occurrent behavior. It is plausible to assume that this will be true of novel media environments as well. Importantly, unlike other forms of media, VR can create a situation in which the user's *entire environment* is determined by the creators of the virtual world, including "social hallucinations" induced by advanced avatar technology. Unlike physical environments, virtual environments can be modified quickly and easily with the goal of influencing behavior.

The comprehensive character of VR plus the potential for the global control of experiential content introduces opportunities for new and especially powerful forms of both mental and behavioral manipulation, especially when commercial, political, religious, or governmental interests are behind the creation and maintenance of the virtual worlds.

However, the plasticity of the mind is not limited to behavioral traits. Illusions of embodiment are possible because the mind is plastic to such a degree that it can misrepresent its own embodiment. To be clear, illusions of embodiment can arise from normal brain activity alone, and need not imply changes in underlying neural structure. Such illusions occur naturally in dreams, phantom limb experiences, out-of-body experiences, and Body Integrity Identity Disorder (Brugger et al., 2000; Metzinger, 2009b; Hilti et al., 2013; Ananthaswamy, 2015; Windt, 2015), and they sometimes include a shift in what has been termed the phenomenal "unit of identification" in consciousness research (UI; Metzinger, 2013a,b), the conscious content that we currently experience as "ourselves" (please note that in the current paper "UI" does not refer to "user interface," but always to the specific experiential content of "selfhood," as explained below). This may be the deepest theoretical reason why we should be cautious about the psychological effects of applied VR: this technology is unique in beginning to target and manipulate the UI in our brain itself.

Direct UI-Manipulation

The UI is the form of experiential content that gives rise to autophenomenological reports of the type "I am this!" For every self-conscious system, there exists a phenomenal unit of identification, such that the system possesses a single, conscious model of reality; the UI is a part of this model; at any given point in time, the UI can be characterized by a specific and determinate representational content, which in turn constitutes the system's phenomenal self-model (PSM, Metzinger, 2003) at *t*. Please note how the UI does not have to be identical with the content of the conscious body image or a region within it (like a fictitious point behind the eyes). For example, the UI can be moved out of and behind the head as phenomenally experienced in a repeatable and controllable fashion by direct electrical stimulation while preserving the visual first-person perspective with its origin behind the eyes (de Ridder et al., 2007). For human beings, the UI is dynamic and can be highly variable. There exists a minimal UI, which likely is constituted by pure spatiotemporal self-location (Blanke and Metzinger,

2009; Windt, 2010; Metzinger, 2013a,b); and in some configurations (e.g., "being one with the world"), there is also a maximal UI, likely constituted by the most general phenomenal property available, namely, the integrated nature of phenomenality *per se* (Metzinger, 2013a,b).

VR technology directly targets the mechanism by which human beings phenomenologically identify with the content of their self-model.

The rubber hand illusion is a simple localized illusion of embodiment that can be induced by having subjects look at a visually realistic rubber hand in a biologically realistic position (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005). When the rubber hand is stroked synchronously with the subject's physical hand (which is hidden from view), subjects experience the rubber hand as their own.³ While the rubber hand can be used to create a partial illusion of embodiment, the same basic idea can be used to create the full-body illusion, on a global level. Subjects look through goggles through which they see a live video feed of their own bodies (or of a virtual body) located a short distance in front of their actual location. When they see their bodies being stroked on the back, and feel themselves being stroked at the same time, subjects sometimes feel as if the body that they see in front of them is their own (Lenggenhager et al., 2007; see **Figure 4**). This illusion is much weaker and more fragile than the RHI, but it has given us valuable new insights into the bottom-up construction of our conscious, bodily self-model in the brain (Metzinger, 2014). In more recent work, Maselli and Slater (2013) have found that tactile feedback is not required for an illusion of embodiment. They found that a virtual arm with a realistic appearance co-located with the subject's actual arm is sufficient to induce the illusion of ownership of the virtual arm. In addition to visual and tactile signals, recent work suggests that manipulations of interoceptive signals, such as heartbeat, can also influence our experience of embodiment (Aspell et al., 2013; Seth, 2013).

The results sketched in these three sections reveal not only categories of risks but also three ways in which the human mind is plastic. First, there is "context-sensitivity all the way down," which may involve hitherto unknown kinds of epigenetic trait formation in new environments. Second, there is evidence that behavior can be strongly influenced by environment and context, and in a deep way. Third, illusions of embodiment can be induced fairly easily in the laboratory, directly targeting the human UI itself. These results can be taken together as empirical premises for an argument stating not only that there may be unexpected psychological risks if illusions of embodiment are misused, or used recklessly, but that, if we are interested in minimizing potential damage and future psychosocial costs, these risks are themselves *ethically* relevant. In the following section, we review initial evidence that connects the three strands of evidence that

³For a version of this illusion using a virtual hand in augmented reality (rather than a rubber hand), see **Figure 2** above.

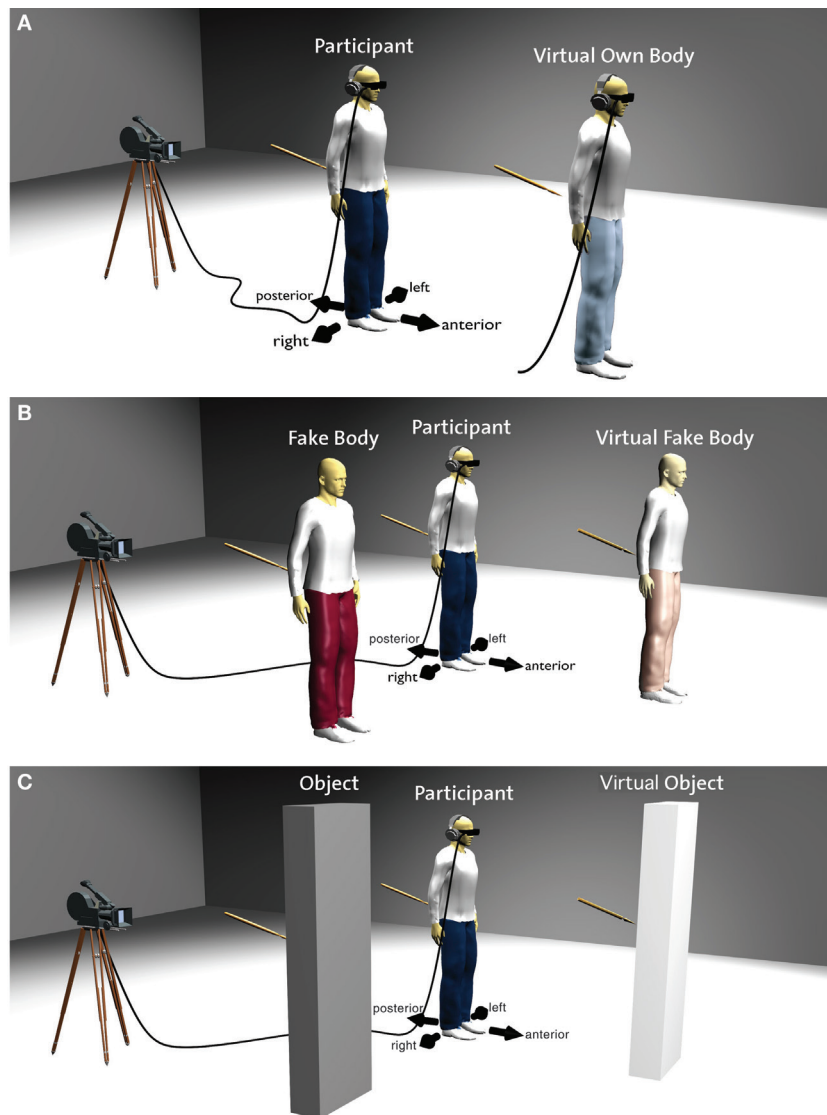


FIGURE 4 | Creating a whole-body analog of the rubber-hand illusion. (A) Participant (dark blue trousers) sees through a HMD his own virtual body (light blue trousers) in 3D, standing 2 m in front of him and being stroked synchronously or asynchronously at the participant's back. In other conditions, the participant sees either **(B)** a virtual fake body (light red trousers) or **(C)** a virtual non-corporeal object (light gray) being stroked synchronously or asynchronously at the back. Dark colors indicate the actual location of the physical body or object, whereas light colors represent the virtual body or object seen on the HMD. (Image used with kind permission from M. Boyer.).

we have just presented. That is, we review initial evidence that illusions of embodiment can be combined with a change in environment and context in order to bring about lasting psychological effects in subjects.

Illusions of Embodiment and Their Lasting Effect

In the last several years, a number of studies have found a psychological influence on subjects while immersed in a virtual environment. These studies suggest that VR poses risks that are novel, that go beyond the risks of traditional psychological experiments in isolated environments, and that go beyond the

risks of existing media technology for the general public. A first important result from VR research involves what is known as the virtual pit (Meehan et al., 2002). Subjects are given a HMD that immerses them in a virtual environment in which they are standing at the edge of a deep pit. In one kind of experiment involving the pit, they are instructed to lean over the edge and drop a beanbag onto a target at the bottom. In order to enhance the illusion of standing at the edge, the subject stands on the ledge of a wooden platform in the lab that is only 1.5" from the ground. Despite their belief that they were in no danger because the pit was "only" virtual, subjects nonetheless show increased signs of stress through increases in heart rate and skin conductance (ibid.). In a

variation of the virtual pit, subjects may be told to walk across the pit over a virtual beam. In the lab, a real wooden beam is placed where subjects see the virtual beam. As one might expect, this version of the pit also elicits strong feelings of stress and fear.⁴ More recently, an experiment reproducing the famous Milgram obedience experiments in VR found that subjects reacted as if the shocks they administered were real, despite believing that they were merely virtual (Slater et al., 2006).

In addition to a strong emotional response from immersion, there is evidence that experiences in VR can also influence behavioral responses. One example of a behavioral influence from VR has been named the *Proteus Effect* by Nick Yee and Jeremy Bailenson. This effect occurs when subjects “conform to the behavior that they believe others would expect them to have” based on the appearance of their avatar (Yee and Bailenson, 2007, p. 274; Kiltani et al., 2013). They found, for example, that subjects embodied in a taller avatar negotiated more aggressively than subjects in a shorter avatar (ibid.). Changes in behavior while in the virtual environment are of ethical concern, since such behavior can have serious implications for our non-virtual physical lives – for example, as financial transactions take place in a non-physical environment (Madary, 2014).

But perhaps even more concerning for our purposes is evidence that behavior while in the virtual environment can have a lasting psychological impact after subjects return to the physical world. Hershfield et al. (2011) found that subjects embodying avatars that look like aged versions of themselves show a tendency to allocate more money for their retirement after leaving the virtual environment. Rosenberg et al. (2013) had subjects perform tasks in a virtual city. Subjects were allowed to fly through the city either using a helicopter or by their own body movements, like Superman. They found that subjects given the superpower were more likely to show altruistic behavior afterwards – they were more likely to help an experimenter pick up spilled pens. Yoon and Vargas (2014) found a similar result, although not using fully immersive VR. They had subjects play a video game as either a superhero, a supervillain, or a neutral control avatar. After playing the game, subjects were given a tasting task that they were told was unrelated to the gaming experiment. Subjects were given either chocolate or chili sauce to taste, and then told to measure out the amount of food for the subsequent subject to taste. Those who played as heroes poured out more chocolate, while those who played as villains poured out more chili.

The psychological impact of immersive VR has also been explored in a beneficent application. Peck et al. (2013) gave subjects an implicit racial bias test at least 3 days before immersion and then immediately after the immersion. In the experiment, subjects were embodied in an avatar with either light skin, dark skin, purple skin, or they were immersed in the virtual world with no body. They found that subjects who were embodied in the dark-skinned avatar showed a decrease in implicit racial bias, at least temporarily.

⁴For some nice anecdotal accounts of experiences with the virtual pit, see Blascovich and Bailenson, 2011: 38–42.

RECOMMENDATIONS FOR THE USE OF VR BY RESEARCHERS AND CONSUMERS

With the results from the first section of the paper in mind as illustrative examples, we now move on to make concrete recommendations for VR in both scientific research (see The Research Ethics of VR) and consumer applications (see Risks for Individuals and Society). Our main recommendations are italicized and listed together in **Table 1**.

The Research Ethics of VR

In this section, we cover questions about the ethics of conducting research either on VR, or, perhaps more interestingly, research *using* VR as a tool. For example, it is plausible to assume that in the future there will be many experiments combining real-time fMRI and VR or ones using animal subjects in VR (Normand et al., 2012), which are not only *about* understanding or improving VR itself but only use it a research tool. To begin with a short example, Behr et al. (2005) have covered the research ethics of VR from a practical perspective, emphasizing that the risk of motion sickness must be minimized and that researchers ought to assist subjects as they leave the virtual environment and readjust to the real world. In this part of the article, we indicate new issues in the research ethics of VR that were not covered in Behr et al.’s initial treatment. In particular, we will raise the following six issues:

- the limits of experimental environments,
- informed consent with regard to the lasting psychological effects of VR,
- risks associated with clinical applications of VR,
- the possibility of using results of VR research for malicious purposes (dual use),
- online research using VR, and
- a general point about the inherent limitations of a code of conduct for research.

For each of these issues, we offer concrete recommendations for researchers using VR as well as ethics committees charged with evaluating the permissibility of particular experimental paradigms using VR.

Ethical Experimentation

What are the limits to what we can do ethically as experiments in VR? We recommend, at the very least, that researchers ought to follow the principle of non-maleficence: do no harm. This principle is a central component of research ethics on human subjects where it is often discussed with the accompanying principle of beneficence: maximize well-being for the subjects. Note how such a principle applies to all sentient beings capable of suffering, like non-human animals or even potential artificial subjects of experience in the future (Althaus et al., 2015, p. 10). We will return to the principle of beneficence in VR in the following section. The principle of non-maleficence can be found in the codes of ethical conduct for both the American Psychological Association (General Principle A)⁵ as well as in the British Psychological

⁵<http://www.apa.org/ethics/code/>

Society (Principle 2.4). The British Psychological Society offers the following recommendation:

Harm to research participants must be avoided. Where risks arise as an unavoidable and integral element of the research, robust risk assessment and management protocols should be developed and complied with. Normally, the risk of harm must be no greater than that encountered in ordinary life, i.e., participants should not be exposed to risks greater than or additional to those to which they are exposed in their normal lifestyles (The British Psychological Society, 2014, p. 11).

Following this recommendation in the case of VR might raise some novel challenges due to the entirely new nature of the technology. For instance, a well-known domain of application for the principle of non-maleficence has been in clinical trials for new pharmacological agents. Although this domain of research ethics still faces important and controversial issues (Wendler, 2012), thinkers in the debate can avail themselves of the history of medical technology. In many cases, precedents can be quoted. In the case of VR, there is yet no history that we can use as a source for insight. On the contrary, what is needed is a rational, ethically sound process of *precedence-setting*.

In its general form, the principle of non-maleficence for VR can be expressed as follows:

No experiment should be conducted using virtual reality with the foreseeable consequence that it will cause serious or lasting harm to a subject.

Although recommending adherence to this principle is nothing new, implementing this principle in VR laboratories may be challenging for the following reason. Attempts to apply non-maleficence in VR can encounter a dilemma of sorts. On the one hand, a goal of the research ought to be, as we suggest below, to gain a better understanding of the risks posed for individuals using VR. For instance, does the duration of immersion pose a greater risk for the user? Might some virtual environments be more psychologically disturbing than others? VR research should seek to answer these and similar questions. In particular, open-ended longitudinal studies will be necessary to assess the risk of long-term usage for the general population, just like with new substances for pharmaceutical cognitive enhancement or medical treatments more generally. On the other hand, it is difficult to assess these risks without running experiments that generate those possible risks, thus raising worries about non-maleficence.

A strict adherence to non-maleficence would require avoiding all experiments using virtual environments for which the risk is unknown. We suggest that this strict interpretation of non-maleficence is not optimal, because substantial ethical assessments should always be evidence-based and necessarily involve the investigation of greater time-windows and larger populations. VR researchers could and should provide a valuable service by informing the public and policy makers of the possible risks of spending large amounts of time in unregulated virtual

environments. The principle of non-maleficence should be applied in the sense that experiments should not be conducted if the outcome involves foreseeable harm to the subjects. On the other hand, the same principle implies a sustained striving for rational, evidence-based minimization of risks in the more distant future. We, therefore, suggest that careful experiments designed with the beneficent intention of discovering the psychological impact of immersion in VR are ethically permissible.

In order to adhere to the principle of non-maleficence, researchers (and ethics committees) will need to utilize their knowledge of experimental psychology as well as their knowledge of results specific to VR. The kinds of results sketched in Sections “Plasticity in the Human Mind” and “Illusions of Embodiment and Their Lasting Effect” will be directly relevant for evaluating whether a line of experimentation violates non-maleficence. Similarly, the selection of subjects for VR experiments must be done with special care. New methods of prescreening for individuals with high risk factors must be incrementally developed, and funding for the development of such new methodologies needs to be allocated. We, therefore, urge careful screening of subjects to minimize the risks of aggravating an existing psychological disorder or an undetected psychiatric vulnerability (Rizzo et al., 1998; Gregg and Tarrier, 2007). Many experiments using VR currently seek to treat existing psychiatric disorders. The screening process for such experiments has the goal of selecting subjects who exhibit signs and symptoms of an existing condition. The screening process should also include *exclusion criteria* specific to possible risks posed by VR. Ideally, the VR research community will seek to establish an empirically motivated standard set of exclusion criteria. As we will discuss in Section “The Effects of Long-Term Immersion” below, of particular concern are vulnerabilities to disorders that could potentially become aggravated by prolonged immersion and illusions of embodiment, such as Depersonalization/Derealization Disorder (DDD; see American Psychiatric Association (2013), DSM-5: 300.14). Standard exclusion criteria may involve, for instance, scoring above a particular threshold on scales testing for dissociative experiences (Bernstein and Putnam, 1986) or depersonalization (Sierra and Berrios, 2000). Of course, there may be cases in which experimenters seek to include subjects with experiences of dissociation in order to investigate ways in which VR might be used to treat the underlying conditions, such as treating post-traumatic stress disorder (PTSD) through exposure therapy in VR (Botella et al., 2015). In those special cases, it is important to implement alternative exclusion criteria, such as Rothbaum et al. (2014), who excluded subjects with a history of psychosis, bipolar disorder, and suicide risk.

Informed Consent

The results presented above clearly suggest that VR experiences can have lasting psychological impact. This new knowledge about the lasting influence of experiences in VR must not be withheld from subjects in new VR experiments.

We recommend that informed consent for VR experiments ought to include an explicit statement to the effect

that immersive VR can have lasting behavioral influences on subjects, and that some of these risks may be presently unknown.

Subjects should be made aware of this possibility out of respect for their autonomy (as included, for example, in the American Psychological Association General Principle D)⁶. That is, if an experiment might alter their behavior without their awareness of this alteration, then such an experiment could be seen as a threat to the autonomy of the subject. A reasonable way to preserve autonomy, we suggest, is simply to inform subjects of possible lasting effects. Please again note the principled problem that research animals are not able to give informed consent, their interest needs to be represented by humans. Also note that we are not suggesting that subjects ought to be informed about the particular effects that are being investigated in the experiment. Thus, our recommendation should not raise the concern that informing subjects may compromise researchers' abilities to test for particular behavioral effects.

Practical Applications: False Hope and Beneficence

Another concern has to do with various applications of VR. One of many promising applications for VR research is in the treatment of disease, damage, and other health-related issues, especially mental health.⁷ For instance, researchers found that immersing burn victims in an icy virtual environment can mitigate their experience of pain during medical procedures (Hoffman et al., 2011). Here, we wish to raise some concerns about applications of VR. The first concern is that patients may develop false hope with regard to clinical applications of VR. The second concern is that applications of VR may encounter a tension between beneficence and autonomy.

Patients may believe that treatment using VR is better than traditional interventions merely due to the fact that it is a new technology, or an experimental application of existing technology. This sense of false hope is known as the "therapeutic misconception" in the literature on the ethics of clinical research (Appelbaum et al., 1987; Kass et al., 1996; Lidz and Appelbaum, 2002; Chen et al., 2003). Researchers using VR for clinical research must be aware of established techniques for combating the therapeutic misconception in their subjects. For example, one established guideline for investigating new clinical applications is that of "clinical equipoise," which is the requirement that there be genuine uncertainty in the medical community as to the best form of treatment (Freedman, 1987). It is important that researchers communicate their own sense of this uncertainty in a clear manner to volunteer subjects. Similarly, as Chen et al. (2003) note, physicians who have a lasting relationship with their patients may be better suited to form a judgment as to whether the patient is motivated by a clear understanding of the nature

of the research, rather than motivated by false hope or even desperation.

VR researchers aiming at new clinical applications should therefore work slowly and carefully, in close collaboration with physicians who may be better situated to make informed judgments about the suitability of particular patients for new trials.

Therapeutic and clinical applications should be investigated only in the presence of certified medical personnel.

Another relevant concern here is the way in which the general public keeps informed of new developments in science through the popular media. Members of the general public with less interest in science may have a more difficult time gleaning scientific knowledge from the media than those with more interest (Takahashi and Tandoc, 2015). When considering their responsibility as scientists to communicate new results to the public (Fischhoff, 2013; Kueffer and Larson, 2014), VR researchers working in clinical applications must be careful to avoid language that might give false hope to patients.

We should also note here that there are other practical concerns about the use of VR for medical interventions. For instance, once the technology is available for patients to use, who will pay for it? Should medical insurance pay for HMDs and new software? How do we achieve distributive justice and avoid a situation where only privileged members of society benefit from technological advances? We make no recommendation here, but flag this question as something that needs to be considered by policy makers. Similarly, HMDs, CAVE immersive displays, and motion-tracking technology may have to be reclassified as medical devices.

One risk when performing the research necessary for developing such applications is that the patients involved may develop a false sense of hope due to the non-traditional nature of the intervention. As this kind of research progresses, scientists must continue to be honest with patients so as not to generate false hope. There is also an overlap between media ethics and the ethics of VR technology: a related example is that many of the early experiments on full-body illusions (Ehrsson, 2007; Lenggenhager et al., 2007) have been falsely overreported as creating full-blown "out-of-body experiences" (Metzinger, 2003, 2009a,b), and scientists have perhaps not done enough to correct this misrepresentation of their own work in the media.⁸ While incremental progress has clearly been made, large parts of the public still falsely believe that scientists "have created OBEs in the lab."

Overall, scientists and the media need to be clear and honest with the public about scientific progress, especially in the area of using VR for medical treatment.

The second concern about applications of VR has to do with the well-known tension between autonomy and beneficence in

⁶<http://www.apa.org/ethics/code/>

⁷VR has been used to treat a wide range of mental health issues, including eating disorders (Ferrer-Garcia et al., 2015), acrophobia (Emmelkamp et al., 2001), agoraphobia (Botella et al., 2004), arachnophobia (Carlin et al., 1997), and PTSD (Rothbaum et al., 2001). See Parsons and Rizzo (2008) for a meta-analysis of these kinds of treatment.

⁸For some examples of the full-body illusion being misrepresented in the media, see: <http://www.nytimes.com/2007/08/24/science/24body.html> <http://news.bbc.co.uk/2/hi/health/6960612.stm> <http://www.sciencedaily.com/releases/2007/08/070823141057.htm>

applied ethics (Beauchamp and Childress, 2013: Chapter 6). As the results surveyed in the first part of this article suggest, VR enables a powerful form of non-invasive psychological manipulation. One obvious application of VR, then, would be to perform such manipulations in order to bring about desirable mental states and behavioral dispositions in subjects. Indeed, early experiments in VR have done just that, making subjects willing to save more for their retirement (Hershfield et al., 2011), perform better on tests for implicit racial bias (Peck et al., 2013), and behave in a more environmentally conscious manner (Ahn et al., 2014). In a paternalistic spirit, such as that of the *UK Behavioral Insights Team*, one might urge that beneficent VR applications such as these should be put in place among the general populace, perhaps as a new form of “public service announcement” for the twenty-first century. Here, we wish to note that doing so may generate another case of conflict between beneficence and autonomy. If individuals do not seek to alter their psychological profile in the ways intended by the beneficent VR interventions, then such interventions may be considered a violation of their autonomy.

Dual Use

Dual use is a well-known problem in research ethics and the ethics of technology, especially in the life sciences (Miller and Selgelid, 2008). Here, we use it to refer to the fact that technology can be used for something other than its intended purpose, in particular to military applications. In the context of VR technology, one will immediately think not only of drone warfare, teleoperated weapon systems, or “virtual suicide attacks,” but also of interrogation procedures and torture. It is not in the power of the scientists and engineers who develop the technology to police its use, but we can raise awareness about potential misuses of the technology as a way of contributing to precautionary steps.

Here is an example. One possible application of VR would be to rehabilitate violent offenders by immersing them in a virtual environment that induces a strong sense of empathy for their victims. We see no problem at all with voluntary participation in such a promising use of the technology. But it is foreseeable that governments and penal systems adopt *mandatory* treatment using similar techniques, calling to mind Anthony Burgess’ *A Clockwork Orange*. We will not comment on the moral acceptability of such a practice, noting that the details of implementation may be an important – and more controllable – unknown factor.

Virtual embodiment constitutes historically new form of acting. Metzinger (2013c) introduced the notion of a “PSM-action” to describe this new element more precisely. PSM-actions are those actions in which a human being exclusively uses the conscious self-model in her brain to initiate an action, causally bypassing the non-neural body (as in **Figure 5**). Of course, there will have to be feedback loops for complex actions, for instance, when seeing through the camera eyes of a robot, perhaps adjusting a grasping movement in real-time (which is still far from possible today). But the relevant causal starting point of the entire action is no longer the body made of flesh and bones, but the conscious self-model in our brain. We simulate an action in the self-model, in the inner *image* of our body, and a machine performs it. PSM-actions are almost purely “mental,” but they may have far-reaching causal consequences in the real world, for example, in combat situations.

As the embodiment in avatars and physical robots may be functionally shallow and may provide only weaker and less stable forms of self-control (for example, with regard to spontaneously arising aggressive fantasies, see Metzinger, 2013c for an example), it is not clear how such PSM-actions mediated via brain-computer interfaces should be assessed in terms of accountability and ethical responsibility.⁹

Just as VR can be used to increase empathy, it can conceivably be used to decrease empathy. Doing so would have obvious military applications in training soldiers to have less empathy for enemy combatants, to feel no remorse about doing violence. We will not go further into the difficult issues regarding the use of new technology in warfare, but we note this possible alternative application of the technology. Apart from increasing or decreasing empathy, the power of VR to induce particular kinds of emotions could be used deliberately to cause suffering. Conceivably, the suffering could be so extreme as to be considered torture. Because of the transparency of the emotional layers in the human self-model (Metzinger, 2003), it will be experienced as real, even if it is accompanied by cognitive-level insight into the nature of the overall situation. Powerful emotional responses occur even when subjects are aware of the fact that they are in a virtual environment (Meehan et al., 2002).

Torture in a virtual environment is still torture. The fact that one’s suffering occurs while one is immersed in a virtual environment does not mitigate the suffering itself.

VR Research and the Internet

A final concern for the research ethics portion of this article has to do with the use of the internet in conjunction with VR research. For instance, scientists may wish to observe the patterns of behavior for users under particular conditions. It is clear that the internet will play a main role in the adoption of VR for personal use. Users will be able to inhabit virtual environments with other users through their internet connections, and perhaps enjoy new forms of avatar-based intersubjectivity. As O’Brocháin et al. (2016) suggest, we will soon see a convergence of VR with online social networks. The overall ethical risks of this imminent development have been covered in detail by O’Brocháin et al. (2016); in this section, we will incorporate and expand on their discussion with a focus on questions of research ethics.

There is a sizable body of literature covering the main issues of internet research ethics (Ess and Association of Internet Researchers Ethics Working Committee, 2002; Buchanan and

⁹It is important to note that teleoperated weapon systems are used in an illegal manner today, and it would not be rational to assume that the introduction of military VR-technology in combination with brain-computer interfaces could lead to a change in this deplorable situation. With German support, the United States of America execute citizens of and in other sovereign states (e.g., Yemen, Somalia, Pakistan) without charge, trial, or final judgment (the so-called “extrajudicial killings”), thereby violating international law (under which lethal force may be used outside armed conflict zones only as a last resort to prevent imminent threats; see Melzer, 2008 for background and discussion) as well as national law, human rights, and humanitarian laws. The potential for further illegal or unethical military applications of VR is high, and one of our major concerns.

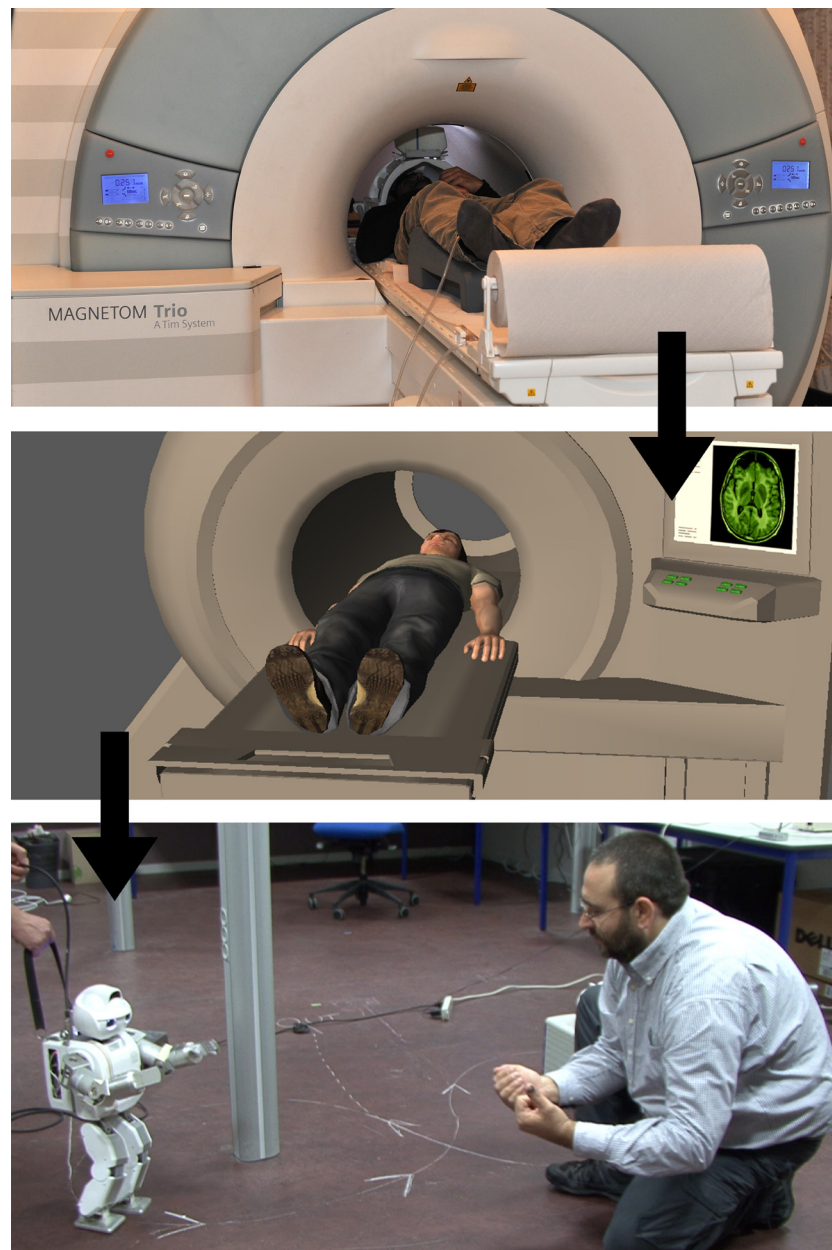


FIGURE 5 | “PSM-actions”: a test subject lies in a nuclear magnetic resonance tomograph at the Weizmann Institute in Israel. With the aid of data goggles he sees an avatar, also lying in a scanner. The goal is to create the illusion that he is embodied in this avatar. The test subject’s motor imagery is classified and translated into movement commands, setting the avatar in motion. After a training phase, test subjects were able to control a remote robot in France “directly with their minds” via the Internet, while they were able to see the environment in France through the robot’s camera eyes. (Image used with kind permission from Doron Friedman and Ori Cohen, cf. Cohen et al., 2014.)

Ess, 2008, 2009). Here, we address the following question: how should these existing issues of internet research ethics be approached for cases of internet research with the use of VR? The two main issues that we will cover here are privacy and obtaining informed consent. We will consider the internet both as a tool and a venue for research (Buchanan and Zimmer, 2015), while noting that virtual environments may place pressure on the distinction between internet as research tool and internet as research venue.

Let us begin with the question of privacy. It is widely accepted that researchers have an ethical obligation to treat confidentially any information that may be used to identify their subjects (see, for example, European Commission, 2013, p. 12). This obligation is based on the general right to privacy outside of a research context (Universal Declaration of Human Rights, article 12, 1948; European Commission Directive 95/46/EC). Practicing this confidentiality may involve, for instance, erasing, or “scrubbing,”

personally identifiable information from a data set (O'Rourke, 2007; Rothstein, 2010).

As O'Brolcháin et al. note, immersive virtual environments will involve the recording of new kinds of personal information, such as “eye-movements, emotions, and real-time reactions” (2015, p. 8). We would like to add that immersive VR could eventually incorporate motion capture technology in order to record the details of users' bodily movements for the purpose of, for example, representing their avatar as moving in a similar fashion. Although implementing this scenario may be beyond the capabilities of the forthcoming commercial hardware, it is plausible and rational to assume that the technology may evolve quickly to include such options. Data regarding the kinematics of users will be useful for researchers from a range of disciplines, especially those interested in embodied cognition (Shapiro, 2014). On the plausible assumption that one's kinematics is very closely related to one's personality and the deep functional structure of bodily self-consciousness – only your body moves in precisely this manner – there will be a highly individual “kinematic fingerprint.” This kind of data collection presents a special threat to privacy.

O'Brolcháin et al. (2016) recommend protecting the privacy for users of online virtual environments through legislation and through incentives to develop new ways of protecting privacy. As a complement to these recommendations, we wish to highlight the threat to privacy created by motion capture technology. Unlike eye-movements and emotional reactions, one's kinematics may be uniquely connected with one's identity, as indicated above. Researchers collecting such data must be aware of its sensitive nature and the dangers of its misuse. In addition, commercial providers of cloud-based VR-technology will frequently have an interest of “harvesting” storing, and analyzing such data and users should be informed about such possibilities and give explicit consent to them.

A second main concern in the ethics of internet research is that of informed consent. In contrast to informed consent for traditional face-to-face experiments, internet researchers may obtain consent by having subjects click “I agree” after being presented with the relevant documentation. There are a number of concerns and challenges regarding the practice of gaining consent for research using the internet as a venue (Buchanan and Zimmer, 2015, see section Privacy, below), including, of course, the fact that actually reading internet privacy policies before accepting them would take far more time than we are willing to allocate – on one estimate, it would take each of us 244 h per year (McDonald and Cranor, 2008).

We suggest that immersive VR will add further complications to these existing issues due to its manipulation of bodily location and its dissolution of boundaries between the real and the virtual. Consider that entering a new internet venue, say a chatroom or a forum, involves a fairly well-defined threshold at which informed consent can be requested before one enters the venue. Due to the centrality of the URL for using the web, one's own location in cyberspace is fairly easy to track. With VR, by contrast, it is foreseeable that one's movement through various virtual environments will be controlled by one's bodily movements, through facial gestures, or simply by the trajectory of visual attention in a way unlike internet navigation using a mouse, keyboard, and

navigation bar. In addition, and more interesting, it is also foreseeable that HMDs will incorporate simultaneous combinations of augmented, substitutional, and VR, with the user being able to toggle between elements of the three. Such a situation would add ambiguity, and perhaps confusion, for attempts to determine the user's location in cyberspace. This ambiguity raises the likelihood that users may give consent for data collection in a particular virtual context but then become unaware of the continued data collection as the user changes context. Such a situation might occur if users of HMDs are able to toggle between, say, an entirely virtual gaming environment, a look out of the window to the busy street below presented through augmented reality, and a family gathering hundreds of kilometers away using substitutional reality through an omni-directional camera set up at the party. This worry can be addressed by giving users continuous reminders (after, of course, they have given informed consent) that their behavior is being recorded for research purposes. Perhaps the visual display could include a small symbol for the duration of the time in which data are being collected.

We leave the implementational details open, but urge the scientific community to take steps to avoid the abuse of informed consent with this technology, especially in the interest of preserving public trust.

A Note on the Limitations of a Code of Ethics for Researchers

We would like to conclude our discussion of the research ethics of VR by noting that the proposed (incomplete) code of conduct is not intended to be sufficient for guaranteeing ethical research in this domain. What we mean here is that following this code should not be considered to be a substitute for ethical reasoning on the part of researchers, reasoning that must always remain sensitive to contextual and implementational details that cannot be captured in a general code of conduct. We urge researchers to conceive of our recommendations here as an aid in their ongoing reflections about the ethical implications and permissibility of their own work, and to proactively support us in developing this ethics code into more detailed future versions. As we emphasized in the beginning of the article, this work is only intended as a first list of possible issues in the research ethics of VR and related technologies. We intend to update and revise this list continuously as new issues arise, although the venue for future revisions is undecided. In any case, we wish to open an invitation for constructive input from researchers in this field regarding issues that should be added or reformulated.

Scientists must understand that following a code of ethics is not the same as being ethical. A domain-specific ethics code, however consistent, developed, and fine grained future versions of it may be, can never function as a substitute for ethical reasoning itself.

Risks for Individuals and Society

Now consider possible issues that may arise with widespread adoption of VR for personal use. Once the technology available

to the general public for entertainment (and other) purposes, individuals will have the option of spending extended periods of time immersed in VR – in a way this is already happening with the advent of smartphones, social networks, increasing time online, etc. Some of the risks and ethical concerns that we have already encountered in the early days of the internet¹⁰ will reappear, though with the added psychological impact enabled by embodiment and a strong sense of presence. We all know that internet technology has long ago begun to change our self-models and consequently our very own psychological structure. The combination with technologies of virtual and robotic re-embodiment may greatly accelerate this development.

For instance, consider the infamous case of virtual rape in LambdaMOO, the text-based multi-user dungeon (MUD). In that virtual world, a player's character known as "Mr.Bungle" used a "voodoo doll" program to control the actions of other characters in the house. He forced them to perform a range of sexual acts, some of which are especially disturbing (Dibbell, 1993). Users of LambdaMOO were outraged, and at least one user whose character was a victim of the virtual rape reported suffering psychological trauma (ibid.). The relevant point to keep in mind here is that this entire virtual transgression occurred in a world that was entirely text based. We will soon be fully immersed in virtual environments, actually embodying – rather than merely describing – our avatars. The results sketched above in Section "Illusions of Embodiment and Their Lasting Effect" suggest that the psychological impact of full immersion will be great, likely far greater than the impact of text-based role-playing. We must now take steps in order to help users avoid suffering psychological trauma of various kinds. To this end, we will discuss four kinds of foreseeable risks:

- long-term immersion;
- neglect of embodied interaction and the physical environment;
- risky content;
- privacy.

We will offer several concrete recommendations for minimizing all four of these kinds of risks to the general public, a number of which call for focused research initiatives.

The Effects of Long-Term Immersion

First, and perhaps most obviously, we simply do not know the psychological impact of long-term immersion. So far, scientific research using VR has involved only brief periods of immersion, typically on the order of minutes rather than hours. Once the technology is adopted for personal use, there will be no limits on the time users choose to spend immersed. Similarly, most research using VR has been conducted using adult subjects. Once VR is available for commercial use, young adults and children will be able to immerse themselves in virtual environments. The risks that we discuss below are especially troublesome for these younger users who are not yet psychologically and neurophysiologically fully developed.

In order to better understand the risks, we recommend longitudinal studies, further research into the psychological effects of long-term immersion.

Of course, these studies must be conducted according to the principles of informed consent, non-maleficence, and beneficence outlined in Section "The Research Ethics of VR." There are several possible risks that can be associated with long-term immersion: addiction, manipulation of agency, unnoticed psychological change, mental illness, and lack of what is sometimes vaguely called "authenticity" (Metzinger and Hildt, 2011, p. 253). The risks that are discovered through longitudinal studies must be directly and clearly communicated to users, preferably communicated *within* VR itself.

Psychologists have long expressed concern about internet use disorder (Young, 1998), and it is a topic of ongoing research (Price, 2011).¹¹ This area of research must now expand in order to include concerns about addiction to immersive VR, both online and offline. Doing so will require monitoring users who prefer to spend long periods of time immersed (see Steinicke and Bruder, 2014 for a first self-experiment). There are two relevant open questions here. First, how might the diagnostic criteria for addiction to VR differ from the established criteria for internet use disorder and related conditions? Note that the neurophysiological underpinnings of VR addiction may differ from that of internet use disorder (Montag and Reuter, 2015) due to the prolonged illusion of embodiment created by VR technology, and because it implies causal interaction with the low-level mechanisms constituting the UI. Second, can we make use of the recommended treatments for internet use disorder for the purpose of helping individuals with VR addiction? For instance, Gresle and Lejoyeux (2011, p. 92) recommend informing users how much time they have spent playing an online game, and including non-player characters in the game to urge players to take breaks. It is plausible that these strategies would be effective for immersive VR as well, but focused research is needed.

A second concern about long-term immersion has to do with the fact that immersive VR can manipulate the user's sense of agency (Gallagher, 2005). In order to generate a strong illusion of ownership for the virtual body, the VR technology must track the self-generated movements of the user's real body and render the virtual body as moving in a similar manner.¹² When things are working well, users experience an illusion of ownership of the virtual body (the avatar is my body), as well as an illusion of agency (I am in control of the avatar). Importantly, the sense of agency in VR is always indirect; control of the avatar is always mediated by the technology. To be more precise, the virtual body representation has been causally coupled with and temporarily embedded into the currently active conscious self-model in the user's brain – it is not that some mysterious "self" leaves the physical body and "enters" the avatar, but rather a novel functional configuration in

¹⁰See Gregory Lastowka, 2010 for a thoughtful treatment of some of the relevant issues.

¹¹Internet use disorder is listed as an area requiring further research in the DSM-5, but it is not (yet) an official disorder according to the manual.

¹²If the real body is not in motion, then co-location of the virtual body with the real body as seen from the first-person perspective can be sufficient for the illusion of ownership (Maselli and Slater, 2013).

which two body representations dynamically interact with each other. However, the causal loop in principle enables *bidirectional* forms of control, or even unnoticed involuntary influence. The fact that the user's sense of agency in VR is always continuously maintained by the technology is an important one for at least two reasons. First, the technology could be used to manipulate users' sense of agency. Second, as we discuss in the general context of mental health below, long-term immersion could cause low-level, initially unnoticeable psychological disturbances involving a loss of the sense of agency for one's physical body.

VR technology could manipulate users' sense of agency by creating a false sense of agency for movements of the avatar that do not correspond to the actual body movements of the user. The same could be true for "social hallucinations," i.e., the creation of the robust subjective impression of ongoing *social* agency, of engaging in a real, embodied form of social interaction, which, however, in reality is only interaction with an unconscious AI or with complex software controlling the simulated social behavior of an avatar. Using only a computer screen, a modified mouse, and headphones, a false sense of agency was created in Daniel Wegner's well-known "I Spy" experiments (Wegner and Wheatley, 1999; Wegner, 2002). In those experiments, subjects reported that they felt themselves to be in control of a cursor selecting an icon on a computer screen when in fact the cursor was being controlled by someone else. The illusion of control was induced by auditory priming – subjects heard a word through headphones that had a semantic association with the icon that was subsequently selected by the cursor. It is reasonable to think that Wegner's method can be implemented rather easily in VR. While immersed in VR, subjects can receive continuous audio and visual cues intended to influence their psychological states. Future experimental work can determine the conditions under which subjects will experience a sense of agency for movements of the avatar that deviate from the subject's actual body movements (as during an OBE or in the dream state, see Kannape et al., 2010 for an empirical study). Important parameters here will likely be the timing of the false movement, the degree to which the false movement deviates from the actual position of the body, and the context of the movement within the virtual environment (including, for instance, the attentional state of the subject).

Creating a false sense of agency in VR is a clear violation of the user's autonomy, a violation that becomes especially worrisome as users spend longer and longer periods of time immersed. Here, we will not insist that all cases of violating autonomy in this manner are ethically impermissible, noting that some such violations may be subtle and beneficent, a kind of virtual "nudge" in the right direction (Thaler and Sunstein, 2009). In addition, human beings often willfully choose to decrease their autonomy, as in drinking alcohol or playing games. But we do claim that creating a false sense of agency in VR is an unacceptable violation of individual autonomy when it is non-beneficent, such as when it is done out of avarice, for example. Manipulating the sense of agency for users in VR is a topic that deserves attention from regulatory agencies.

A third concern that we wish to raise about long-term immersion is that of risks for mental health. As stated above, we simply

do not know whether long-term immersion poses a threat for mental health. Future research ought to investigate whether factors such as the duration of immersion, the content of the virtual environment (including the user's own avatar or the way in which the software controls the automatic behavior, facial gestures, or gaze of *other* avatars), and the user's pre-existing psychological profile might have lasting negative effects on the mental health of users. As mentioned above (see Ethical Experimentation), we suspect that heavy use of VR might trigger symptoms associated with Depersonalization/Derealization Disorder (DSM-5 300.14). Overall, the disorder can be characterized as having chronic feelings or sensations of unreality. In the case of depersonalization, individuals experience an unreality of the bodily self, and in the case of derealization, individuals experience the external world as unreal. For instance, those suffering from the disorder report feeling as if they are automata (loss of the sense of agency), and feeling as if they are living in a dream (see Simeon and Abugel, 2009 for illustrative reports from individuals suffering from depersonalization).¹³ Note that Depersonalization/Derealization Disorder involves feelings of unreality but not delusions of unreality, there is a dissociation of the low-level phenomenology of "realness" from high-level cognition. That is, someone suffering from depersonalization may lose the sense of agency, but will not thereby form the false belief that they are no longer in control of their own actions.

Depersonalization/Derealization Disorder is relevant for us here because VR technology manipulates the psychological mechanisms involved in generating experiences of "realness," mechanisms similar or identical to those that go awry for those suffering from the disorder. Even though users of VR do not believe that the virtual environment is real, or that their avatar's body is really their own, the technology is effective because it generates illusory feelings as if the virtual world is real (recall the virtual pit from Section "Illusions of Embodiment and Their Lasting Effect" above). What counts is the variable degree of transparency or opacity of the user's own conscious representations (Metzinger, 2003). Our concern is that long-term immersion could cause damage to the neural mechanisms that create the feeling of reality, of being in immediate contact with the world and one's own body. Heavy users of VR may begin to experience the real world and their real bodies as unreal, effectively shifting their sense of reality exclusively to the virtual environment.¹⁴ We recommend focused longitudinal studies on the impact on mental health of long-term immersion in VR. These studies should especially investigate risks for dissociative disorders, such as Depersonalization/Derealization Disorder.

¹³There is a sizeable literature on depersonalization/derealization. Some of the central works include Steinberg and Schnall, 2001; Radovic and Radovic, 2002; Simeon and Abugel, 2009; Sierra, 2012.

¹⁴We should be clear here that we are only speculating about a possible causal connection between long-term immersion and experiences of depersonalization/derealization. The etiology of the disorder is still not well understood. It is well-known that episodes of depersonalization/derealization can be triggered by stress, panic attacks, and the use of some drugs (Simeon, 2004). One prominent theory suggests that chronic depersonalization/derealization may be caused by childhood trauma (ibid.), though see Marshall et al. (2000).

A final concern for long-term immersion stems from the fact that some may consider experiences in the virtual environment to be “inauthentic,” because those experiences are artificially generated. This concern may remind some readers of Robert Nozick’s well-known thought experiment about an “Experience Machine” that can provide users with any experience they desire (Nozick, 1974, p. 42–45). Nozick uses the thought experiment to raise a problem for utilitarianism, urging his readers to consider reasons why one might not wish to “plug-in” to the machine, claiming that “something matters to us in addition to experience” (Nozick, 1974, p. 44). The interesting question, of course, now becomes what would happen if this “additional something” can be added to the experience itself, for example by advanced VR-technology creating the *phenomenal* quality of “authenticity,” of direct reference to something “meaningful,” for example by a more robust version of naïve realism on the level of subjective experience itself or by manipulation of the user’s emotional self-model. While Nozick suggests that many of us would not wish to plug-in to the experience machine for the reason just stated, recent work by Felipe de Brigard suggests otherwise. De Brigard (2010) presented students with several variations on the thought experiment all with an important twist on Nozick’s original version. In de Brigard’s version, we are told that we are already plugged-in to an experience machine and we are asked if we would like to unplug in order to return to our “real” lives. Many of de Brigard’s students replied that they would not wish to unplug, leading de Brigard to suggest that our reactions to the thought experiment are influenced more by the *status quo* bias (Samuelson and Zeckhauser, 1988) than by our valuing of something more than experience. It is the *status quo* bias, de Brigard suggests, that gives us pause about plugging-in to the machine (in Nozick’s version of the thought experiment) just as it is the *status quo* bias that gives us pause about unplugging (in de Brigard’s version).

Overall, de Brigard’s results offer initial reasons to be skeptical about Nozick’s supposition that we would not plug-in because we value factors beyond experience alone. Even with this skepticism, though, many of us may still feel that there is something false, “inauthentic,” or undesirable about living large portions of one’s life in an entirely artificial environment, such as VR. Apart from the dubious essentialist metaphysics lurking behind the vague and sometimes ideologically charged notion of an “authentic self,” it is important to note how such intuitions are historically plastic and culturally embedded: they may well change over time as larger parts of the population begin to use advanced forms of VR technology. As an example, please note how already today we find a considerable number of people who are not able to grasp the difference between “friendship” and “friendship on Facebook” any more. Fully engaging with the issue of losing “authenticity” in virtual environments would likely require entering into some deep philosophical waters, and we are unable to do so here, though we will touch on some of the relevant issues below. Apart from the deeper philosophical issues, there is one important point that we wish to make before moving on.

The point has to do with the way in which we imagine the possibilities of VR for personal use. One reason behind an assertion that long-term immersion would be an inauthentic way of spending one’s time is that one might assume that the content of

immersive VR would be unedifying, making people more shallow as they retreat from society in favor of an artificial social world in which their decisions are made for them. Brolcháin et al. raise this sort of concern:

With little exposure to “higher” culture, to great works of art and literature; and without the skills (and maybe the attention spans) to enjoy them; people would be less able to engage with the world at a deep level. People without exposure to great works and ideas might find that [their] inner lives are shaped to a large degree by market-led cultural products rather than works of depth and profundity. (2015, p. 20)

We agree that such a scenario would be undesirable, but wish to counterbalance this concern by reminding readers that it is not unique to VR. It is a concern that can be applied in various degrees to other media technology as well, going all the way back to worries about the written word in Plato (*Phaedrus* 274d–275e). The printing press, for example, can enable one to disseminate great works of literature, but it can also enable the dissemination of vulgarity – and it certainly changed our minds. Readers with vulgar tastes can “immerse” themselves as they wish. The same goes for photography and motion pictures. The important point is as follows. There is no reason to doubt that works of great depth and profundity can be produced by artists who choose VR as their medium. Just as film emerged as a new predominant art form in the twentieth century, so might VR in the twenty-first century. We predict that immersive VR-technology will gradually lead to the emergence of completely new forms of art (or even architecture, see Pasqualini et al., 2013), which may be hard to conceive today, but which will certainly have cultural consequences, perhaps even in our understanding what an artistic subject and esthetic subjectivity really are.

Neglect of Others and the Physical Environment

As users spend increasing time in virtual environments, there is also a risk of their neglecting their own bodies and physical environments – just as for many people today posing and engaging in disembodied social interactions via their Facebook account has become more important than what was called “real life” in the past. In extreme cases, individuals refuse to leave their homes for extended periods of time, behavior categorized as “Hikikomori” by the Japanese Ministry of Health. VR will enable us to interact with each other in new ways, not through disembodied interaction, as in the texts, images, and videos of current social media, but rather through what we have called the illusion of embodiment. We will interact with other avatars while embodied in our own avatars. Or perhaps we will use augmented reality through omni-directional cameras that allow us to enjoy the illusion of being in the presence of someone who is far away in space and/or time. To put it more provocatively, we may soon, as Norbert Wiener anticipated many years ago, have the ability to “telegraph” human beings (Wiener, 1954, p. 103–104). Telepresence is likely to become a much more accessible, immediate, comprehensive, and embodied experience.

Our general recommendation on this theme is for focused research into the following question: What, if anything, is lost in cases of social interactions that are mediated using advanced telepresence in VR? If such losses were unnoticed, what negative effects for the human self-model could be expected?

This question has been a major theme in some of Hubert Dreyfus' work on the philosophy of technology. Dreyfus has emphasized that mediating technologies may not capture something of what is important for real-time interactions in the flesh, what, following Merleau-Ponty, he calls "intercorporeality" (Dreyfus, 2001, p. 57). When we are not present in the flesh with others, the context and mood of a situation may be difficult to appreciate – if only because the bandwidth and the resolution of our internal models are much lower. Perhaps more importantly, there is a concern that mediating technologies will not allow us to pick up on all of the subtle bodily cues that appear to play a major role in social communication through unconscious entrainment (Frith and Frith, 2007), cues that involve ongoing embodied interaction (Gallagher, 2008; de Jaegher et al., 2010).

In addition to the concerns about losing embodied signaling for communication, we might also consider what is lost from the sense modalities that are not (yet) integrated into VR. As Sherry Turkle puts it, when these kinds of technology "keep grandparents from making several-thousand-mile treks to see their grandchildren in person (and there is already evidence that they do), children will be denied something precious: the starchy feel of a grandmother's apron, the smell of her perfume up close, and the taste of her cooking" (Harmon, 2008; Turkle, 2011, p. 342). Advances in technology could conceivably address Turkle's point about other perceptual modalities, but there remains a question about what may be lost even if we can create virtual content for other sense modalities.¹⁵ One recent finding that should raise concern here is that depression is more likely in older adults who have less social contact in person regardless of their amount of telephone, written, and email contact (Teo et al., 2015). Apart from this troubling finding, even if the technology eventually enables rich social interaction through telepresence, the concern remains that heavy use of such technology will lead to neglect or even animosity toward one's actual physical and social environment. The recurring tragedies of parents with "gamer rage" who have injured and killed their children because

the children disrupt their playing indicate that this concern is valid and serious.¹⁶

Clark (2003) takes a notably different approach to these kinds of issues, raising the point that, instead of treating VR and related technologies as a replacement for in-the-flesh interaction, we should think of them as providing opportunities for new and perhaps *enhanced* modes of human interaction. Rather than unsatisfactory reproductions of familiar modes of interaction, the technology should be developed with an eye toward "expanding and reinventing our sense of body and action" (2003, p. 111). Consider, for example, using a combination of substitutional and augmented reality to see a representation of some of the physiological states of your partner who is many miles away – such as a soft flash over the body in synchrony with the heartbeat (as in Aspell et al., 2013). That and similar uses of the technology could plausibly enhance embodied (though mediated) social interaction. As with many other topics addressed here, future research will be crucial for our understanding of which uses of the technology will be best for enabling positive forms of (mediated) social interaction.

Clark's recommendation that we use the technology as an enhancer rather than a replacement does have some appeal. However, what counts as an "enhancement," and what as therapy or mere life-style decision, has been a topic of ethical debates for a long time, for instance, in assessing the correct use of pharmaceutical cognitive enhancement (Metzinger and Hildt, 2011; Metzinger, 2012). We should also note that his recommendation may not entirely address the concerns raised by Dreyfus, Turkle, and others. The foreseeable problem is that the general public simply will not share Clark's vision, choosing to use the technology as a *de facto* replacement for traditional modes of interaction (as Turkle notes in the passage above). Are "Facebook-friendships" social enhancements or social disabilities? In such a situation, we must remain mindful of what may be lost, especially when the technology may encourage less frequent "in-the-flesh" visits to the infirm and immobile.

We wish to close this discussion of the ways in which VR might attenuate our contact with others and with our physical environments by revisiting a point briefly made in the previous section on a loss of authenticity during long-term immersion. As noted above in the discussion of Nozick's experience machine, many readers might have the intuition that spending long periods of time in virtual environments is "somehow inauthentic." Yet, what counts for the applied ethics of VR are not intuitions, but rather rational arguments and empirical evidence. We would like to note that a likely relevant factor here may be whether those long periods of immersion involve forms of intersubjective engagement with others that are subjectively experienced as meaningful, and how this experience is integrated into our culture. Along these lines, one may suggest that the artificial nature of the virtual environment is not as important compared to whether or not the environment affords intersubjective engagement experienced as meaningful (see Bostrom, 2003, p. 245–55 and Chalmers, 2005

¹⁵All of these concerns bring up the question of whether the problem is merely a shortcoming in the technology, or something more fundamental. That is, should we only be concerned about losing important information through mediated interactions, information such as bodily cues and tactile sensations? If so, then advances in technology can conceivably address that concern. Or is there something else that is lost when not present in the flesh with others? It seems that thinkers such as Dreyfus wish to suggest that there is something else that is lost when we lose "intercorporeality," something that cannot be captured with better and better technology. Still, it remains somewhat difficult to articulate what that "something else" might be. One possibility is that social interactions that are mediated by advanced technology lose some form of "authenticity" as discussed above. It is also worth noting that our epistemic limitations may be relevant: in the case of VR, we do not yet know the way in which social interaction will be altered.

¹⁶For a list of examples, see: <http://movingtolearn.ca/2013/gamer-rage-child-abuse-a-growing-problem-deserving-our-attention> (retrieved 1 December 2015).

for similar ideas). This, of course, opens the possibility that ultimately shallow or even largely meaningless social interactions (once again, think of today's Facebook-“friendships” and “likes”) are *experienced* as substantial by users who are really only overwhelmed by the possibilities of future VR-technology, and which are subsequently *described* as meaningful. A shallow form of social interaction could then become culturally assimilated and thereby “normalized” (Metzinger and Hildt, 2011, p. 247). Normalization is a complex sociocultural process by which certain new norms become accepted in societal practice, a process that is often mediated by the availability of new technologies, a process that changes our very own minds and which, therefore, carries the risk of unnoticed self-deception. Here, we cannot explore this rich (and controversial) philosophical territory, but note that it may be relevant for grappling with the worry of “inauthenticity” in virtual environments.

Risky Content

Another main concern for users of VR is that of virtual content. One might begin with the general rule of thumb that red lines not to be crossed in reality should be the default red lines in VR. One obvious problem, though, is that users will almost certainly seek out VR as a way of crossing red lines with impunity. A second possible problem is that this rule of thumb would make VR even more subjectively real. One main issue here is whether some particular kinds of content in VR should be discouraged in various ways. Obvious candidates for such content would be sex (virtual pedophilia, virtual rape) and violence. But there are perhaps less obvious kinds of content that should be considered, such as content encouraging and reinforcing undesirable personality traits, including those identified as the “dark triad” (Paulus and Williams, 2002). The dark triad refers to narcissism, Machiavellianism, and psychopathy. Individuals may find it appealing to spend time in virtual worlds designed to reward characters that exhibit traits associated with the dark triad. For example, the MMORPG *EVE Online* is known for fostering a style of play that involves manipulating and deceiving other players. The VR version of *EVE Online*, *EVE: Valkyrie*, has been described as “[u]ndoubtedly the most heavily anticipated virtual reality game.”¹⁷ Based on some of the empirical results surveyed above (see “Illusions of Embodiment and Their Lasting Effect”), there is cause for concern about behavioral patterns rewarded in immersive games such as *EVE: Valkyrie* having a lasting influence on the psychological profile of users.

Apart from the behaviors encouraged by particular virtual environments, there are concerns about the content that can be created when users will have the freedom to create and design their own avatars. For instance, one goal of our own project VERE is to create software that enables untrained users to generate an avatar that resembles any human being with fairly little time and effort. This application would in principle allow for “body swapping,” in which users enter the bodies of others (Petkova and Ehrsson, 2008). It is also worth noting that these avatars will be

available for use after their human model is dead. Thus, we will be able to “resurrect” the dead in VR. The ability to body-swap and to interact with the dead in this way may offer great opportunity for therapy in the hands of the beneficent, but it could easily lead to profound trauma, especially in the hands of characters such as Mr. Bungle, mentioned above.

These considerations raise difficult questions about which regulatory actions would facilitate the best overall outcome. On the one hand, there are good reasons for taking a fairly restrictive approach to avatar ownership. On the other hand, there are also reasons for allowing individuals maximum freedom in their creation and use of avatars. Of course, one's approach to such questions will likely reflect whether one's political philosophy has more paternalistic or libertarian leanings. We will consider the reasons for each approach in turn.

A reasonable starting point on this issue would be to treat avatars in an analogous manner to personality rights relating to the publication of photos. They are public representations of persons. Interestingly, societies and legal systems exhibit considerable differences in their underlying moral intuitions here. One important conceptual issue here may be *determining the relevant degree of similarity* between an avatar and a human person. Just as many accept the right of an individual to control the commercial use of his or her name, image, likeness, one might, for example, interpret the “right to my own avatar” a property right as opposed to a personal right. Therefore, the validity of the right of publicity could be taken to survive the death of the biological individual. There will be new questions about the ownership (and individuation) of avatars. The likeness between a person and their avatar may or may not be an important factor. Instead of likeness, we might individuate avatars by a unique proper name that can be represented in the virtual space, as in many video games. How does one assign an unequivocal identity to the virtual representation of a body or a person? Could there be something like a chassis plate number, a license plate, or a “virtual vehicle identification number? (VVIN)? We already have digital object identifiers (DOIs) for electronic documents and other forms of content, a form of persistent identification, with the goal of permanently and unambiguously identifying the object with which a given DOI is associated. But what about an avatar that is currently *used* by a human operator, namely by functionally and phenomenologically identifying with it? Should we dynamically associate a “digital subject identifier” (DSI) with it? There will also be questions about whether some kinds of virtual activities should be censored. Examples of such activities having to do with sex and violence are left to the reader's imagination. Another kind of content worth considering may be the use of virtual environments for indoctrination into extremist groups.

With these initial thoughts in place, now consider the reasons for taking a fairly strict regulatory stance on the ownership of one's own avatar. After mentioning the pressure from social networks such as Facebook for users to use their real identities, O'Brolcháin et al. recommend the development of technologies similar to digital watermarking that would ensure “that only the genuine owner of an avatar can use it” (2015, p. 22). Would this perhaps have to be a “DSI,” as we proposed above? They suggest

¹⁷<http://www.craveonline.com/culture/878953-top-10-virtual-reality-games-will-convince-strap-vr-headset#/slide/10> (retrieved 30 September 2015).

that such technology would help protect the autonomy and privacy of users. Along the same lines, if one were to identify strongly with one's own avatar, the "theft" and use of that avatar by another may be extremely disturbing. Importantly, avatar theft may also create completely new opportunities for impersonation and fraud, for example also using physical robots.

From a more theoretical point of view, we might distinguish between internal and external self-models: the internal self-model is in the brain of the user, and it is grounded in his or her body (Metzinger, 2014), whereas external person- and body-models can be created in virtual environments. Here, the specific, historically new kind of action that needs to be ethically assessed and legally regulated takes place when a user *identifies* with a potential external model of the self by dynamically integrating it with the internal model of the self already active in his or her brain. The core question seems to be what consequences we draw from the potential for phenomenological ownership to legal notions of ownership. Virtual identification can cause real suffering, and real suffering is relevant for the law.

Without denying the value of protecting avatar ownership, we would now like to consider two reasons for taking a less restrictive, more libertarian, approach. First, implementing control over the use of particular avatars may be impractical. So far, attempts to curb digital piracy using technology have not been very successful, and there is no reason to think that things will be different for avatars. In fact, regulation and control may be even more difficult with avatars due to questions raised above having to do with avatar individuation and degrees of similarity. Say a user creates an avatar that is similar but not pixel-for-pixel identical to another user's avatar.¹⁸ Where precisely should we draw the line between theft and acceptable similarity? Protecting avatar ownership might lead to a regulatory quagmire. Even if the appearance of the avatar is not highly relevant for ownership, we would need to establish a widely accepted alternative method of individuation, such as a unique proper name that cannot be easily forged. The second reason for taking a less restrictive approach would be out of concern for individual creative freedom. As noted above, VR holds the promise of being a powerful new artistic medium – the creative possibilities are astonishing. The fact that regulations on avatar ownership may restrict those possibilities must be taken into consideration.

Avatar ownership and individuation will be an important issue for regulatory agencies to consider. There are strong reasons to place restrictions on the way in which avatars can be used, such as protecting the interests and privacy of individuals who strongly identify with their own particular avatar on social networks. On the other hand, these restrictions may prove impractical to implement and may unnecessarily limit personal creative freedom.

¹⁸The importance of personal identity for moral philosophy is well-known (Parfit, 1984; Shoemaker, 2014). The considerations here introduce the additional complication of identity for virtual representations of persons. See Vallor (2010, especially pp. 166–167) and Rodogno (2012) for insightful discussions.

Privacy

Privacy is, of course, a major concern with contemporary information technology (van den Hoven et al., 2014), and there are further concerns about privacy with the foreseeable convergence between VR and social networks (O'Brolcháin et al., 2016). Here, we wish to offer only a few quick remarks on this topic, noting that this issue deserves further attention. Commercial applications of virtual environments introduce new possibilities for targeted advertising or "neuromarketing," thus attacking the individual's mental autonomy (Metzinger, 2015). By tracking the details of one's movements in VR, including eye movements, involuntary facial gestures, and other indicators of what researchers call low-level intentions or "motor intentions" (Riva et al., 2011), private agencies will be able to acquire details about one's interests and preferences in completely new ways (Coyle and Thorson, 2001). If avatars themselves should in the future be used as "humanoid interfaces," consumers can be influenced and manipulated by real-time feedback of the avatar's own facial and eye movements (for example, via automatic and unconscious responses in their mirror-neuron system; Rizzolatti and Craighero, 2004). Commercials in VR could even feature images of the target audience himself or herself using the product. The use of big data to "nudge" users ("Big Nudging") combined with VR could have long-lasting effects, perhaps producing changes in users' mental mechanisms themselves.

Users ought to be made aware that there is evidence that advertising tactics using embodiment technology such as VR can have a powerful unconscious influence on behavior.

SUMMARY

In this article, we have considered some of the risks that may arise with the commercial and research use of VR. We have offered some concrete recommendations and noted areas in which further ethical deliberation will be required. One main theme of the article is that there are several open empirical questions that should be urgently addressed in a beneficent research environment in order to mitigate risks and raise awareness for users of VR in the general public. More research is needed. Here, one of our main goals was to provide a first set of ethical recommendations as a platform for future discussions, a set of normative starting points that can be continuously refined and expanded as we go along (see **Table 1**).

Let us end by making one more general point, an observation which is of a more philosophical nature. VR is the representation of *possible* worlds and *possible* selves, with the aim of making them appear as real as possible – ideally, by creating a subjective sense of "presence" in the user. Interestingly, some of our best theories of the human mind and conscious experience itself describe it in a very similar way: leading current theories of brain dynamics (Friston, 2010; Hohwy, 2013; Clark, 2015) describe it as the constant creation of internal *models* of the world, predictively generating hypotheses – virtual neural representations – about

TABLE 1 | VERE code of conduct for the ethical use of VR in research and by the general public.

RECOMMENDATIONS FOR THE RESEARCH ETHICS OF VR	
1. Non-maleficence	<ul style="list-style-type: none"> a. No experiment should be conducted using virtual reality with the foreseeable consequence that it will cause involuntary suffering or serious or lasting harm to a subject. b. A rational, evidence-based identification and minimization of risks (also those pertaining to a more distant future) ought to be a part of research itself.
2. Informed consent	<ul style="list-style-type: none"> a. Informed consent for VR experiments ought to include an explicit statement to the effect that immersive VR can have lasting behavioral influences on subjects, and that some of these risks may be presently unknown. b. Experimental VR research should not be carried out on subjects incapable of informed consent.
3. Transparency and media ethics	<ul style="list-style-type: none"> a. In experimental work developing new clinical applications, researchers should be careful not to create false hopes in patients by repeatedly reminding them of the merely experimental nature of the research. b. VR researchers aiming at new clinical applications should work in close collaboration with physicians who may be better situated to make informed judgments about the suitability of particular patients for new trials. c. Scientists and the media need to be clear and honest with the public about scientific progress, and not only in the area of using VR for medical treatment. d. In interacting with the media, scientists should cultivate a proactive attitude, especially if they are the first to become aware of novel types of risks through their own work. Communication with the public, if needed, should be self-initiated, an act of taking control and acting in advance of a future situation, rather than just reacting.
4. Dual use	<ul style="list-style-type: none"> a. Potential military applications of VR, AR, and SR should be closely monitored by policy makers and funding agencies alike. b. Torture in a virtual environment is still torture. The fact that one's suffering occurs while one is immersed in a virtual environment does not mitigate the suffering itself. c. Policy makers should aim at international arrangements among countries to add VR, AR, and SR in a process to harmonize lists of dual-use technologies to be controlled.
5. Internet research	<ul style="list-style-type: none"> a. The scientific community has to take steps to avoid the abuse of informed consent with this technology, especially in the interest of preserving public trust. b. The ability to toggle between VR, AR, and SR may create situations in which users are not able to maintain an understanding of when their informed consent to share information is in effect. Users should be repeatedly reminded within VR that they have given informed consent.
6. The Limitations of a Code of Conduct	<ul style="list-style-type: none"> a. Scientists must understand that following a code of ethics is not the same as <i>being</i> ethical. A domain-specific ethics code, however consistent, developed, and fine-grained future versions of it may be, can never function as a substitute for ethical reasoning itself. b. Such reasoning must be conducted in a way that is sensitive to the contextual and implementational details of particular experimental paradigms, details that cannot be captured by a general code of conduct.
RECOMMENDATIONS FOR THE USE OF VR BY THE GENERAL PUBLIC	
1. Long-term immersion	<ul style="list-style-type: none"> a. Longitudinal studies and further research into the psychological effects of long-term immersion are needed. b. Users must be made aware that these studies are seriously limited in that they will, due to ethical constraints, exclude users who may be most vulnerable (such as children or those with latent mental illness). Some of these vulnerabilities may be unknown to science and unknown to the users themselves.
2. Increasing virtualization of social interactions – we call for focused research, large longitudinal studies, into the following questions:	<ul style="list-style-type: none"> a. What, if anything, is lost in cases of social interactions that are mediated using advanced telepresence in VR? b. If such losses were unnoticed, what negative effects for the human self-model could be expected?
3. Risky content	<ul style="list-style-type: none"> a. As compared to the viewing of traditional movies containing graphic violence or pornography, the impact of full immersion settings and the associated risk of users suffering psychological trauma will steadily increase as VR technology advances. Users have to be made aware of this possibility. b. VR technology holds the potential to create robust social hallucinations, to directly manipulate the sense of agency, to modulate personality traits via identification with virtual characters, or to causally interact with deeper levels of self-consciousness (UI-manipulation). Users have to be made aware of this possibility. c. Avatar ownership will be an important issue for regulatory agencies to consider. There are strong reasons to place restrictions on the way in which avatars can be used, such as protecting the interests and privacy of individuals who strongly identify with their own particular avatar on social networks. On the other hand, these restrictions may prove impractical to implement and may unnecessarily limit personal creative freedom. Regulators must strike a rational balance between these concerns.
4. Privacy	<ul style="list-style-type: none"> a. Users ought to be made aware that there is evidence that advertising tactics using embodiment technology, such as VR, can have a powerful unconscious influence on behavior. For example, a combination of “Big Nudging” strategies (collecting big data for the purposes of nudging the general public) with VR technology could have long-lasting effects, which might also affect underlying mental mechanisms themselves. b. Data protection: users ought to be made aware of new risks involving surveillance, such as reading out “motor intentions” or a “kinematic fingerprint” during avatar use.

the hidden causes of sensory input through probabilistic inference. Slightly earlier, some philosophers (Revonsuo, 1995, p. 55; Revonsuo, 2009, p. 115.; Metzinger, 2003, p. 556; 2008, 2009a, p. 6, 23, pp. 104–108; Westerhoff, 2015 for overview and discussion) have pointed out at length how conscious experience exactly is a virtual model of the world, a dynamic internal simulation, which in standard situations cannot be experienced as a virtual model because it is phenomenally transparent – we “look through it” as if we were in direct and immediate contact with reality. What is historically new, and what creates not only novel psychological risks but also entirely new ethical and legal dimensions, is that *one* VR gets ever more deeply embedded into *another* VR: the conscious mind of human beings, which has evolved under very specific conditions and over millions of years, now gets causally coupled and informationally woven into technical systems for representing possible realities. Increasingly, it is not only culturally and socially embedded but also shaped by a technological niche that over time itself quickly acquires a rapid, autonomous dynamics and ever new properties. This creates a complex convolution, a nested form of information flow in which the biological mind and its technological niche influence each other in ways we

are just beginning to understand. It is this complex convolution that makes it so important to think about the Ethics of VR in a critical, evidence-based, and rational manner.

AUTHOR CONTRIBUTIONS

MM drafted the initial outline of the article and then both MM and TM contributed written content.

ACKNOWLEDGMENTS

The authors wish to thank Mel Slater, Patrick Haggard, Angelika Peer, and all members of the VERE project for critical discussion and detailed proposals. We are also grateful to the three reviewers of this contribution for recommending substantial improvements to the original draft.

FUNDING

This work was funded by FP7 257695 Virtual Embodiment and Robotic Re-Embodiment (VERE).

REFERENCES

- Ahn, S. J., Bailenson, J., and Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Comput. Human Behav.* 39, 235–245. doi:10.1016/j.chb.2014.07.025
- Althaus, D., Erhardt, J., Gloor, L., Hutter, A., Mannino, A., and Metzinger, T. (2015). “Künstliche Intelligenz: Chancen und Risiken,” in *Diskussionspapiere der Stiftung für Effektiven Altruismus*, Vol. 2, 1–17. Available at: <http://ea-stiftung.org/s/Kunstliche-Intelligenz-Chancen-und-Risiken.pdf>
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders. DSM-5*, 5th Edn. Washington, DC: Author.
- Ananthaswamy, A. (2015). *The Man Who Wasn't There. Investigations into the Strange New Science of the Self*. London: Dutton.
- Appelbaum, P. S., Roth, L. H., Lidz, C. W., Benson, P., and Winslade, W. (1987). False hopes and best data: consent to research and the therapeutic misconception. *Hastings Cent. Rep.* 17, 20–24. doi:10.2307/3562038
- Asch, S. (1951). “Effects of group pressure upon the modification and distortion of judgment,” in *Groups, Leadership and Men: Research in Human Relations*, ed. H. Guetzkow (Oxford: Carnegie Press), 177–190.
- Aspell, J., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., and Blanke, O. (2013). Turning body and self inside out: visualized heartbeats alter bodily self-consciousness and tactile perception. *Psychol. Sci.* 24, 2445–2453. doi:10.1177/0956797613498395
- Azuma, R. (1997). A survey of augmented reality. *Presence* 6, 355–385. doi:10.1162/pres.1997.6.4.355
- Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* 2, 412–414. doi:10.1098/rsbl.2006.0509
- Beauchamp, T., and Childress, J. (2013). *Principles of Biomedical Ethics*, 7th Edn. New York: Oxford University Press.
- Behr, K.-M., Nosper, A., Klimmt, C., and Hartmann, T. (2005). Some practical considerations of ethical issues in VR research. *Presence* 14, 668–676. doi:10.1162/105474605775196535
- Bernstein, E. M., and Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *J. Nerv. Ment. Dis.* 174, 727–735. doi:10.1097/00005053-198612000-00004
- Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi:10.1016/j.tics.2008.10.003
- Blascovich, J., and Bailenson, J. (2011). *Infinite Reality. Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution*, 1st Edn. New York: William Morrow.
- Borgmann, A. (1984). *Technology and the Character of Contemporary life. A Philosophical Inquiry*. Chicago: University of Chicago Press.
- Bostrom, N. (2003). Are we living in a computer simulation? *Philos. Q.* 53, 243–255. doi:10.1111/1467-9213.00309
- Botella, C., Serrano, B., Baños, R., and Garcia-Palacios, A. (2015). Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: a review of its efficacy, the adequacy of the treatment protocol, and its acceptability. *Neuropsychiatr. Dis. Treat.* 11, 2533–2545. doi:10.2147/NDT.S89542
- Botella, C., Villa, H., García Palacios, A., Quero, S., Baños, R., and Alcaniz, M. (2004). The use of VR in the treatment of panic disorders and agoraphobia. *Stud. Health Technol. Inform.* 99, 73–90.
- Botvinick, M., and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature* 391, 756. doi:10.1038/35784
- Brey, P. (2010). Philosophy of technology after the empirical turn. *Tech. Res. Philos. Technol.* 14, 36–48. doi:10.5840/techne20101416
- Brugger, P., Kollias, S. S., Müri, R. M., Crelrier, G., Hepp-Reymond, M. C., and Regard, M. (2000). Beyond re-membering: phantom sensations of congenitally absent limbs. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6167–6172. doi:10.1073/pnas.100510697
- Bublitz, J. C., and Merkel, R. (2014). Crimes against minds: on mental manipulations, harms and a human right to mental self-determination. *Crim. Law Philos.* 8, 51–77. doi:10.1007/s11572-012-9172-y
- Buchanan, E., and Ess, C. (2008). “Internet research ethics. The field and its critical issues,” in *The Handbook of Information and Computer Ethics*, eds K. Himma and H. Tivani (Hoboken, NJ: Wiley), 273–292.
- Buchanan, E., and Zimmer, M. (2015). “Internet research ethics,” in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Available at: <http://plato.stanford.edu/archives/spr2015/entries/ethics-internet-research/>
- Buchanan, E. A., and Ess, C. M. (2009). Internet research ethics and the institutional review board. *SIGCAS Comput. Soc.* 39, 43–49. doi:10.1145/1713066.1713069
- Carlin, A. S., Hoffman, H. G., and Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behav. Res. Ther.* 35, 153–158. doi:10.1016/S0005-7967(96)00085-X
- Chalmers, D. (2005). “The matrix as metaphysics,” in *Philosophers Explore the Matrix*, ed. C. Grau (Oxford, UK: Oxford University Press), 132–176.

- Chen, D., Miller, F., and Rosenstein, D. (2003). Clinical research and the physician-patient relationship. *Ann. Int. Med.* 138, 669–672. doi:10.7326/0003-4819-138-8-200304150-00015
- Clark, A. (2003). *Natural-Born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*. Oxford, NY: Oxford University Press.
- Clark, A. (2015). *Surfing Uncertainty. Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Cohen, O., Koppel, M., Malach, R., and Friedman, D. (2014). Controlling an avatar by thought using real-time fMRI. *J. Neural Eng.* 11, 035006. doi:10.1088/1741-2560/11/3/035006
- Coyle, J., and Thorson, E. (2001). The effects of progressive levels of interactivity and vividness in web marketing sites. *J. Advert.* 30, 65–77. doi:10.1080/00913367.2001.10673646
- de Brigard, F. (2010). If you like it, does it matter if it's real? *Phil. Psychol.* 23, 43–57. doi:10.1080/09515080903532290
- de Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi:10.1016/j.tics.2010.06.009
- de Ridder, D., van Laere, K., Dupont, P., Menovsky, T., and Van de Heyning, P. (2007). Visualizing out-of-body experience in the brain. *N. Engl. J. Med.* 357, 1829–1833. doi:10.1056/NEJMoa070010
- Dibbell, J. (1993). “A rape in cyberspace,” in *The Village Voice*. Available at: <http://www.villagevoice.com/news/a-rape-in-cyberspace-6401665>
- Dreyfus, H. (2001). *On the Internet*. Routledge.
- Ehrsson, H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi:10.1126/science.1142175
- Emmelkamp, P. M., Bruynzeel, M., Drost, L., and van der Mast, C. A. (2001). Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *Cyberpsychol. Behav.* 4, 335–339. doi:10.1089/109493101300210222
- Ess, C., and Association of Internet Researchers Ethics Working Committee. (2002). *Ethical Decision-Making and Internet Research*. Recommendations from the AoIR Ethics Working Committee; Association of Internet Researchers. Available at: <http://aoir.org/reports/ethics.pdf>
- European Commission. (2013). *Ethics for Researchers. Facilitating Research Excellence in FP7*. Brussels. Available at: http://ec.europa.eu/research/participants/data/ref/fp7/89888/ethics-for-researchers_en.pdf
- Fan, K., Izumi, H., Sugiura, Y., Minamizawa, K., Wakisaka, S., Inami, M., et al. (2013). “Reality jockey: lifting the barrier between alternate realities through audio and haptic feedback,” in *SIGCHI Conference*, eds W. Mackay, S. Brewster, and S. Bødker (Paris: ACM SIGCHI), 2557.
- Ferrer-Garcia, M., Gutierrez-Maldonado, J., Treasure, J., and Vilalta-Abella, F. (2015). Craving for food in virtual reality scenarios in non-clinical sample: analysis of its relationship with body mass index and eating disorder symptoms. *Eur. Eat. Disord. Rev.* 23, 371–378. doi:10.1002/erv.2375
- Fischhoff, B. (2013). The sciences of science communication. *Proc. Natl. Acad. Sci. U.S.A.* 110(Suppl. 3), 14033–14039. doi:10.1073/pnas.1213273110
- Franssen, M., Lokhorst, G.-J., and van de Poel, I. (2009). “Philosophy of technology,” in *Stanford Encyclopedia of Philosophy*, ed. Z. Edward. Available at: <http://plato.stanford.edu/entries/technology/>
- Freedman, B. (1987). Equipoise and the ethics of clinical research. *N. Engl. J. Med.* 317, 141–145. doi:10.1056/NEJM198707163170304
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi:10.1038/nrn2787
- Frith, C., and Frith, U. (2007). Social cognition in humans. *Curr. Biol.* 17, R724–R732. doi:10.1016/j.cub.2007.05.068
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford, NY: Clarendon Press.
- Gallagher, S. (2008). Inference or interaction: social cognition without precursors. *Phil. Explor.* 11, 163–174. doi:10.1080/13869790802239227
- Gregg, L., and Tarrier, N. (2007). Virtual reality in mental health. A review of the literature. *Soc. Psychiatry Psychiatr. Epidemiol.* 42, 343–354. doi:10.1007/s00127-007-0173-4
- Gregory Lastowka, F. (2010). *Virtual Justice. The New Laws of Online Worlds*. New Haven, CT: Yale University Press.
- Gresle, C., and Lejoyeux, M. (2011). “Phenomenology of internet addiction,” in *Internet Addiction*, ed. O. P. Hannah (Hauppauge, NY: Nova Science Publisher's, Inc), 85–94.
- Haney, C., Banks, W., and Zimbardo, P. (1973). Study of prisoners and guards in a simulated prison. *Nav. Res. Rev.* 9, 1–17.
- Harmon, A. (2008). “Grandma's on the computer screen,” in *The New York Times*. Available at: <http://www.nytimes.com/2008/11/27/us/27minicam.html>
- Heeter, C. (1992). Being there: the subjective experience of presence. *Presence* 1, 262–271. doi:10.1162/pres.1992.1.2.262
- Heidegger, M. (1977). *The Question Concerning Technology, and Other Essays*, 1st Edn. New York: Harper & Row (Harper colophon books).
- Hershfield, H., Goldstein, D., Sharpe, W., Fox, J., Yeykelis, L., Carstensen, L., et al. (2011). Increasing saving behavior through age-progressed renderings of the future self. *J. Mark. Res.* 48, S23–S37. doi:10.1509/jmkr.48.SPL.S23
- Hilti, L., Hänggi, J., Vitacco, D., Kraemer, B., Palla, A., Luechinger, R., et al. (2013). The desire for healthy limb amputation: structural brain correlates and clinical features of xenomelia. *Brain* 136(Pt 1), 318–329. doi:10.1093/brain/aw316
- Hoffman, H., Chambers, G., Meyer, W., Arceneaux, L., Russell, W., Seibel, E., et al. (2011). Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *Ann. Behav. Med.* 41, 183–191. doi:10.1007/s12160-010-9248-7
- Hohwy, J. (2013). *The Predictive Mind*, First Edn. Oxford: Oxford University Press.
- Huang, Z., Hui, P., Peylo, C., and Chatzopoulos, D. (2013). Mobile augmented reality survey: a bottom-up approach. *The Computing Research Repository (CoRR)*. Available at: <http://arxiv.org/abs/1309.4413>.
- Kannape, O. A., Schwabe, L., Tadi, T., and Blanke, O. (2010). The limits of agency in walking humans. *Neuropsychologia* 48, 1628–1636. doi:10.1016/j.neuropsychologia.2010.02.005
- Kass, N. E., Sugarman, J., Faden, R., and Schoch-Spana, M. (1996). Trust, The fragile foundation of contemporary biomedical research. *Hastings Cent. Rep.* 26, 25–29. doi:10.2307/3528467
- Kilteni, K., Bergstrom, I., and Slater, M. (2013). Drumming in immersive virtual reality: the body shapes the way we play. *IEEE Trans. Vis. Comput. Graph.* 19, 597–605. doi:10.1109/TVCG.2013.29
- Kueffer, C., and Larson, B. M. H. (2014). Responsible use of language in scientific writing and science communication. *Bioscience* 64, 719–724. doi:10.1093/biosci/biu084
- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi:10.1126/science.1143439
- Lidz, C., and Appelbaum, P. (2002). The therapeutic misconception: problems and solutions. *Med. Care* 40(9 Suppl.), V55–V63. doi:10.1097/01.MLR.0000023956.25813.18
- Madary, M. (2014). Intentionality and virtual objects: the case of Qiu Chengwei's dragon sabre. *Ethics Inf. Technol.* 16, 219–225. doi:10.1007/s10676-014-9347-4
- Marcuse, H. (1964). *One-Dimensional Man. Studies in the Ideology of Advanced Industrial Society*, 2nd Edn. London: Routledge.
- Marshall, R. D., Schaefer, F. R., Lin, S. H., Simpson, H. B., Vermes, D., and Liebowitz, M. (2000). Childhood trauma and dissociative symptoms in panic disorder. *Am. J. Psychiatry* 157, 451–453. doi:10.1176/appi.ajp.157.3.451
- Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7:83. doi:10.3389/fnhum.2013.00083
- McDonald, A., and Cranor, L. F. (2008). The cost of reading privacy policies. *I/S J. Law Policy Inform. Soc.* 4.
- Meehan, M., Insko, B., Whitton, M., and Brooks, F. (2002). Physiological measures of presence in stressful virtual environments. *ACM Trans. Graph.* 21, 645–652. doi:10.1145/566654.566630
- Melzer, N. (2008). *Targeted Killing in International Law*. Oxford, NY: Oxford University Press (Oxford Monographs in International Law).
- Metz, R. (2012). Augmented reality is finally getting real. *MIT Technol. Rev.* 2.
- Metzinger, T. (2003). *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Metzinger, T. (2008). Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples. *Prog. Brain Res.* 168, 215–278. doi:10.1016/S0079-6123(07)68018-2
- Metzinger, T. (2009a). *The Ego Tunnel. The Science of the Mind and the Myth of the Self*. New York: Basic Books.
- Metzinger, T. (2009b). Why are out-of-body experiences interesting for philosophers? The theoretical relevance of OBE research. *Cortex* 45, 256–258. doi:10.1016/j.cortex.2008.09.004
- Metzinger, T. (2012). Zehn Jahre Neuroethik des pharmazeutischen kognitiven Enhancements – aktuelle Probleme und handlungsrichtlinien für die Praxis. *Fortschr. Neurol. Psychiatr.* 80, 36–43. doi:10.1055/s-0031-1282051

- Metzinger, T. (2013c). "Two principles for robot ethics," in *Robotik und Gesetzgebung*, eds E. Hilgendorf and J.-P. Günther (Baden-Baden: Nomos), 247–286.
- Metzinger, T. (2013a). The myth of cognitive agency: subpersonal thinking as a cyclically recurring loss of mental autonomy. *Front. Psychol.* 4:931. doi:10.3389/fpsyg.2013.00931
- Metzinger, T. (2013b). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Front. Psychol.* 4:746. doi:10.3389/fpsyg.2013.00746
- Metzinger, T. (2014). "First-order embodiment, second-order embodiment, third-order embodiment," in *The Routledge Handbook of Embodied Cognition*, ed. Lawrence A. S. (London: Routledge), 272–286.
- Metzinger, T. (2015). M-autonomy. *J. Conscious Stud.* 22, 11–12.
- Metzinger, T., and Hildt, E. (2011). "Cognitive enhancement," in *The Oxford Handbook of Neuroethics*, eds J. Illes and B. J. Sahakian (Oxford, NY: Oxford University Press (Oxford Library of Psychology)), 245–264.
- Milgram, P., and Colquhoun, H. (1999). "A taxonomy of real and virtual world display integration," in *Mixed Reality: Merging Real and Virtual Worlds*, eds Y. Ohta and H. Tamura (New York: Springer), 5–30.
- Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* E77-D, 1321–1329.
- Milgram, S. (1974). *Obedience to Authority. An Experimental View*. London: Tavistock.
- Miller, S., and Selgelid, M. (2008). *Ethical and Philosophical Consideration of the Dual-Use Dilemma in the Biological Sciences*. New York: Springer.
- Montag, C., and Reuter, M. (2015). *Internet Addiction. Neuroscientific Approaches and Therapeutical Interventions (Studies in Neuroscience, Psychology and Behavioral Economics)*. New York: Springer.
- Normand, J.-M., Sanchez-Vives, M., Waechter, C., Giannopoulos, E., Grosswindhager, B., Spanlang, B., et al. (2012). Beaming into the rat world: enabling real-time interaction between rat and human each at their own scale. *PLoS one* 7:e48331. doi:10.1371/journal.pone.0048331
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- O'Brolcháin, F., Jacquemard, T., Monaghan, D., O'Connor, N., Novitzky, P., and Gordijn, B. (2016). The convergence of virtual reality and social networks: threats to privacy and autonomy. *Sci. Eng. Ethics* 22, 1–29. doi:10.1007/s11948-014-9621-1
- O'Rourke, P. (2007). *Report of the Public Responsibility in Medicine and Research. Human Tissue/Specimen Banking Working Group*. Available at: <http://www.primr.org/workarea/downloadasset.aspx?id=936>
- Osimo, S., Pizarro, R., Spanlang, B., and Slater, M. (2015). Conversations between self and self as Sigmund Freud – a virtual body ownership paradigm for self counselling. *Sci. Rep.* 5, 13899. doi:10.1038/srep13899
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Parsons, T., and Rizzo, A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *J. Behav. Ther. Exp. Psychiatry* 39, 250–261. doi:10.1016/j.jbtep.2007.07.007
- Pasqualini, I., Llobera, J., and Blanke, O. (2013). "Seeing" and "feeling" architecture: how bodily self-consciousness alters architectonic experience and affects the perception of interiors. *Front. Psychol.* 4:354. doi:10.3389/fpsyg.2013.00354
- Paulus, D., and Williams, K. (2002). The dark triad of personality. Narcissism, Machiavellianism, and psychopathy. *J. Res. Pers.* 36, 556–563. doi:10.1016/S0092-6566(02)00505-6
- Peck, T., Seinfeld, S., Aglioti, S., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi:10.1016/j.concog.2013.04.016
- Petkova, V., and Ehrsson, H. (2008). If i were you: perceptual illusion of body swapping. *PLoS ONE* 3:e3832. doi:10.1371/journal.pone.0003832
- Price, H. (ed.) (2011). *Internet Addiction*. Hauppauge, NY: Nova Science Publisher's, Inc.
- Radovic, F., and Radovic, S. (2002). Feelings of unreality: a conceptual and phenomenological analysis of the language of depersonalization. *Phil. Psychiatr. Psychol.* 9, 271–279. doi:10.1353/ppp.2003.0048
- Revonsuo, A. (1995). Consciousness, dreams and virtual realities. *Phil. Psychol.* 8, 35–58. doi:10.1080/09515089508573144
- Revonsuo, A. (2009). *Inner Presence. Consciousness as a Biological Phenomenon*. Cambridge, MA: MIT Press.
- Riva, G., Waterworth, J., Waterworth, E., and Mantovani, F. (2011). From intention to action: the role of presence. *New Ideas Psychol.* 29, 24–37. doi:10.1016/j.newideapsych.2009.11.002
- Rizzo, A. A., Wiederhold, M., and Buckwalter, J. G. (1998). Basic issues in the use of virtual environments for mental health applications. *Stud. Health Technol. Inform.* 58, 21–42.
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi:10.1146/annurev.neuro.27.070203.144230
- Rodogno, R. (2012). Personal identity online. *Philos. Technol.* 25, 309–328. doi:10.1007/s13347-011-0020-0
- Rosenberg, R., Baughman, S., and Bailenson, J. (2013). Virtual superheroes: using superpowers in virtual reality to encourage prosocial behavior. *PLoS ONE* 8:e55003. doi:10.1371/journal.pone.0055003
- Rothbaum, B., Price, M., Jovanovic, T., Norrholm, S., Gerardi, M., Dunlop, B., et al. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder in Iraq and Afghanistan war veterans. *Am. J. Psychiatry* 171, 640–648. doi:10.1176/appi.ajp.2014.13121625
- Rothbaum, B. O., Hodges, L. F., Ready, D., Graap, K., and Alarcon, R. D. (2001). Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. *J. Clin. Psychiatry* 62, 617–622. doi:10.4088/JCP.v62n0808
- Rothstein, M. (2010). Is deidentification sufficient to protect health privacy in research? *Am. J. Bioethics* 10, 3–11. doi:10.1080/15265161.2010.494215
- Samuelson, W., and Zeckhauser, R. (1988). Status quo bias in decision making. *J. Risk Uncertainty* 1, 7–59. doi:10.1007/bf00055564
- Seth, A. (2013). Interceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi:10.1016/j.tics.2013.09.007
- Shapiro, L. (2014). *The Routledge Handbook of Embodied Cognition*, 1 Edn. London: Routledge, Taylor & Francis Group (Routledge handbooks).
- Shoemaker, D. (2014). "Personal Identity and Ethics," in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Available at: <http://plato.stanford.edu/archives/spr2014/entries/identity-ethics/>
- Sierra, M. (2012). *Depersonalization. A New Look at a Neglected Syndrome*. Cambridge: Cambridge Medicine.
- Sierra, M., and Berrios, G. E. (2000). The Cambridge depersonalization scale: a new instrument for the measurement of depersonalization. *Psychiatry Res.* 93, 153–164. doi:10.1016/S0165-1781(00)00100-1
- Simeon, D. (2004). Depersonalisation disorder: a contemporary overview. *CNS Drugs* 18, 343–354. doi:10.2165/00023210-200418060-00002
- Simeon, D., and Abel, J. (2009). *Feeling Unreal. Depersonalization Disorder and the Loss of the Self*. Oxford, NY: Oxford University Press.
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., et al. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* 1:e39. doi:10.1371/journal.pone.0000039
- Slater, M., Spanlang, B., Sanchez-Vives, M., and Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE* 5:e10564. doi:10.1371/journal.pone.0010564
- Steinberg, M., and Schnall, M. (2001). *The Stranger in the Mirror. Dissociation: the Hidden Epidemic*. New York: Cliff Street Books.
- Steinicke, F., and Bruder, G. (2014). "A self-experimentation report about long-term use of fully-immersive technology," in *The 2nd ACM Symposium*, eds W. Andy, S. Frank, S. Evan, and S. Wolfgang (Honolulu, HI: ACM), 66–69.
- Suzuki, K., Wakisaka, S., and Fujii, N. (2012). Substitutional reality system: a novel experimental platform for experiencing alternative reality. *Sci. Rep.* 2, 459. doi:10.1038/srep00459
- Takahashi, B., and Tandoc, E. (2015). Media sources, credibility, and perceptions of science: learning about how people learn about science. *Public Underst. Sci.* doi:10.1177/0963662515574986
- Teo, A., Choi, H., Andrea, S., Valenstein, M., Newsom, J., Dobscha, S., et al. (2015). Does mode of contact with different types of social relationships predict depression in older adults? Evidence from a nationally representative survey. *J. Am. Geriatr. Soc.* 63, 2014–2022. doi:10.1111/jgs.13667
- Thaler, R., and Sunstein, C. (2009). *Nudge. Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin Books.
- The British Psychological Society. (2014). *Code of Human Research Ethics*. Available at: http://www.bps.org.uk/system/files/Public%20files/code_of_human_research_ethics_dec_2014_inf180_web.pdf
- Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91. doi:10.1037/0096-1523.31.1.80

- Turkle, S. (2011). *Alone Together. Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.
- Vallor, S. (2010). Social networking technology and the virtues. *Ethics Inf. Technol.* 12, 157–170. doi:10.1007/s10676-009-9202-1
- van den Hoven, J., Blaauw, M., Pieters, W., and Warnier, M. (2014). “Privacy and information technology,” in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Available at: <http://plato.stanford.edu/archives/win2014/entries/it-privacy/>
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wegner, D. M., and Wheatley, T. (1999). Apparent mental causation. Sources of the experience of will. *Am. Psychol.* 54, 480–492. doi:10.1037/0003-066X.54.7.480
- Wendler, D. (2012). “The ethics of clinical research,” in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Available at: <http://plato.stanford.edu/archives/fall2012/entries/clinical-research/>
- Westerhoff, J. (2015). What it means to live in a virtual world generated by our brain. *Erkenn.* 1–22. doi:10.1007/s10670-015-9752-z
- Wiener, N. (1954). *The Human Use of Human Beings. Cybernetics and Society*. New York, NY: Da Capo Press (The Da Capo series in science).
- Windt, J. (2010). The immersive spatiotemporal hallucination model of dreaming. *Phenom. Cogn. Sci.* 9, 295–316. doi:10.1007/s11097-010-9163-1
- Windt, J. (2015). *Dreaming. A Conceptual Framework for Philosophy of Mind and Empirical Research*. Cambridge, MA: MIT Press.
- Yee, N., and Bailenson, J. (2007). The proteus effect. The effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x
- Yoon, G., and Vargas, P. (2014). Know thy avatar: the unintended effect of virtual-self representation on behavior. *Psychol. Sci.* 25, 1043–1045. doi:10.1177/0956797613519271
- Young, K. (1998). Internet addiction: the emergence of a new clinical disorder. *Cyberpsychol. Behav.* 1, 237–244. doi:10.1089/cpb.1998.1.237

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Madary and Metzinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.