



Real moral problems in the use of virtual reality

Erick Jose Ramirez¹ · Scott LaBarge¹

Published online: 26 July 2018
© Springer Nature B.V. 2018

Abstract

In this paper, we argue that, under a specific set of circumstances, designing and employing certain kinds of virtual reality (VR) experiences can be unethical. After a general discussion of simulations and their ethical context, we begin our argument by distinguishing between the experiences generated by different media (text, film, computer game simulation, and VR simulation), and argue that VR experiences offer an unprecedented degree of what we call “perspectival fidelity” that prior modes of simulation lack. Additionally, we argue that when VR experiences couple this perspectival fidelity with what we call “context realism,” VR experiences have the ability to produce “virtually real experiences.” We claim that virtually real experiences generate ethical issues for VR technologies that are unique to the medium. Because subjects of these experiences treat them as if they were real, a higher degree of ethical scrutiny should be applied to any VR scenario with the potential to generate virtually real experiences. To mitigate this unique moral hazard, we propose and defend what we call “The Equivalence Principle.” This principle states that “if it would be wrong to allow subjects to have a certain experience in reality, then it would be wrong to allow subjects to have that experience in a virtually real setting.” We argue that such a principle, although limited in scope, should be part of the risk analysis conducted by any Institutional Review Boards, psychologists, empirically oriented philosophers, or game designers who are using VR technology in their work.

Keywords Applied ethics · Institutional review boards · Media experience · Moral psychology · Phenomenology · Virtual Reality

Virtual reality (VR) technologies¹ are generating significant interest among philosophers and psychologists, in part because this technology might help us to examine the nature of morality and moral decision-making through observation of agents’ moral choices in VR scenarios. Additionally, psychological experiments using VR have become more frequent and complex over the last 10 years, while simultaneously video game developers and other creators are using VR to offer us experiences in a wide range of new imagined places and settings. But along with these exciting possibilities come potential hazards. In this paper, we argue that, under a specific set of circumstances, designing and employing certain kinds of VR experiences can be unethical.

After a general discussion of simulations and their ethical context, we begin our argument by distinguishing between the experiences generated by different media (text, film, computer game simulation, and VR simulation), and argue that VR experiences offer an unprecedented degree of what we will call “perspectival fidelity” that prior modes of simulation lack. Additionally, we argue that when VR experiences couple this perspectival fidelity with what we will call “context-realism,” VR technology has an unparalleled ability to produce “virtually real experiences.” We claim that virtually real experiences generate ethical issues for VR developers that are uniquely pressing for the medium. Because subjects of virtually real experiences treat them *as if they were real*, a higher degree of ethical scrutiny should

✉ Erick Jose Ramirez
ejramirez@scu.edu

Scott LaBarge
slabarge@scu.edu

¹ Philosophy Department, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, USA

¹ VR technologies comprise a large class of hardware devices that can include room-sized projection systems into which subjects are placed, head-mounted displays, and augmented reality (AR) devices which overlay additional content onto a subject’s experience of the actual world (Parsons et al. 2017). We focus on head-mounted displays because such systems are, by far, the most widespread form of VR researchers and the public are likely to use. Although we focus our analysis on head-mounted VR displays, much of what we say will also apply to other forms of VR and AR interfaces.

be applied to any VR scenario with the potential to generate virtually real experiences.

At the end of our argument we offer a heuristic that we refer to as “The Equivalence Principle (TEP).” This principle, while it does not apply uniquely to virtual reality technologies, states that “if it would be wrong to allow subjects to have a certain experience in reality, then it would be wrong to allow subjects to have that experience in a virtually real setting.” We argue that such a heuristic, although limited in scope, can help to reframe the evaluation of VR simulations by Institutional Review Boards, psychologists, empirically oriented philosophers, or game designers who are using VR technology in their work.

On simulation

Let us begin with the question, what constitutes a simulation? In asking this question, we set foot in well-trod territory involving issues like the nature of theoretical models, but we do not intend to take a strong stand in that area. For our purposes, it is not a matter of great importance whether simulations and models are the same thing or not. We are content for people to use whatever theory of simulations and models they like, so long as they count the relevant sorts of examples as simulations. In this paper, we consider a simulation to be anything which attempts to reproduce another thing in such a way that the simulation is relevantly like the simulated thing without being exactly like it.² An implication of this definition is that there must be some significant way in which the simulation is unlike the thing it simulates; if there were no such difference between them, the simulation would be instead a true reproduction, recreation, or reinstatement.³

² The substantial philosophical literature on simulation is centered heavily around the role of simulations and models in experimental science, and the metaphysical and epistemological issues that are prominent in that discussion are not particularly germane to our concerns in this paper. However, we share some areas of overlapping interest. Frigg and Hartmann (2012) have a useful discussion of a range of views concerning what has to be true of a model for it to successfully “represent” its target. On the difference between models and simulations, see Krohs (2008) and Morrison (2009). Winsberg (2009) distinguishes some different sorts of simulation. Knuttila (2011) has a useful discussion of the senses in which scientific models may be said to represent the physical reality they model that could have some bearing on simulations and the things they simulate, particularly emphasizing the intentions of the creators/users of the simulation. Godfrey-Smith (2006, p. 733) points out how different scientists can construe the same model as having different success criteria in a way that tracks our point below concerning the context-dependent nature of success in simulation.

³ For example, the computer program *Microsoft Flight Simulator*, in simulating what it is like to fly an airplane, attempts to be relevantly like an actual airplane by providing the user of the simulation with visual and auditory feedbacks that are experientially similar to what

One key aspect of our definition requires that a simulation be “relevantly like” the thing simulated. What this means in any given case depends a lot on context and the intentions of the creators. For any simulated object, the creators of the simulation might choose any number of features the object possesses to capture with their simulation. For instance, programmers who are creating a computer flight simulator might concentrate on different features of the experience of flying a plane. They might seek to simulate the physics of airplane flight particularly faithfully without worrying much about creating photo-realistic versions of the cockpit and the world outside, or they might make the opposite choice and emphasize photo-realism over a close approximation of real-world physics. Or, of course, they might choose to capture both, given sufficient processing power, knowledge, and skill. Presumably there are any number of other features of the experience of airplane flight they might choose to simulate. The essential point is that different simulations of the same thing can be relevantly like the thing simulated in very different ways.

Similarly, what counts as success for a simulation depends on what features the simulation is trying to capture in the simulated object. For example, a weather model successfully predicts the weather when the variables it tracks interact in a way that predicts the actual weather accurately. Because such a model does not attempt to simulate what it is like to witness a storm, it would be a mistake to criticize the model for not reproducing the sound of a predicted storm’s thunder. By contrast, if a film aims to simulate the experience of a thunderstorm, it is perfectly reasonable to judge the simulation’s success by the extent to which the film makes its viewers feel like they are experiencing a storm, even if the depiction of the weather renders the real-world science of storms inaccurately in all sorts of ways. Like the film of a storm, many simulations attempt to recreate an *experience* of their simulated object as faithfully as possible, and in such cases we can judge the simulations’ success or failure by that degree of faithfulness to the original, but where the target of

Footnote 3 (continued)

an actual pilot would hear and see in her airplane (e.g. the clouds and horizon, the instrument panel, the roar of the engines), but it remains a simulation because there is no actual airplane involved, and the game “pilot” never leaves her desk chair. If one were to try to simulate flying an airplane by putting someone in an actual airplane and having them work the real controls to really fly the airplane, that person wouldn’t be *simulating* flying the plane, she would be actually flying it. On this view, while a digital environment could simulate a real environment, it could never be a reinstatement or reproduction of that environment. It would lack the necessary substance. It may nevertheless be possible, however, that certain *elements* of the digital environment may be reinstatements of *elements* of the real environment. The example of reproduced sounds that we discuss later in the paper would constitute such a case.

the simulation is something other than experiential fidelity, we will need to set our standards of success differently.

Preliminary remarks on ethics in simulation

When we consider the ethics of simulation, we need to differentiate several aspects of the simulation process that require distinct ethical attention. On the one hand, we need to consider the interests and responsibilities of the different agents involved. First, we should identify those agents who *experience* the simulation, i.e. the subjects or users of the simulation. How does the simulation affect them? Is there any way in which the simulation might be said to meaningfully harm them? Second, we need to consider possible indirect impacts the simulation might have on other agents whom the subjects of the simulation might *interact with* later, outside of the simulation itself. For instance, if users of a video game simulation are exposed in the course of the simulation to racist or sexist attitudes or practices, it is at least conceivable that those users might treat women or racial minorities differently when they are not playing the game. Moreover, whatever the impact of the simulation on specific players, the simulation might maintain and normalize pernicious social attitudes. Of course, a lot of concern about video games, films, and other simulation media has focused on these aspects of the ethics of simulation. In these ways, even if a simulation does not directly harm its subjects themselves, it might be a contributing cause of harm to others who never experienced the simulation in virtue of encouraging users to simulate immoral actions.⁴ And finally, we need to consider the *creators* of the simulation, who bear at least some of the responsibility for any foreseeable harm that their simulation might cause.

On the other hand, keeping in mind the roles of these agents, we also need to consider different areas of concern with simulations themselves. First, we need to evaluate the *content* of a simulation, since any number of ethical issues may arise from that content. Many examples of ethically problematic content come quickly to mind: we might be concerned about violent or pornographic content, or about harmful stereotypes of various groups that might be encoded in a simulation. Likewise, if the content of a simulation portrays its simulated object inaccurately or conveys misinformation, it could contribute to seriously harmful errors of

⁴ We thank an anonymous reviewer for pressing us on this issue and for helping us clarify the nature of our particular moral concern. Although we believe that encouraging users to simulate immoral actions raises concerns about the nature of simulation, our own concerns in this article focus on what we believe is a different and under-examined problem concerning the possibility of VR experience itself to cause subjects harm.

various sorts. Indeed, a lot of the moral complaints about various forms of simulative media have focused on these sorts of content-based issues.⁵

A different sort of concern about simulations, however, focuses on the *medium* in which the simulation is made. Different sorts of simulative media have different characteristic features, and the users of these media experience the relevant simulations in sometimes very different ways. Content that might be unproblematic when simulated in one medium might conceivably become problematic when simulated in a medium with different features. While we are not uninterested in the ethical issues that arise from the content of simulations, we will not address them here. Instead, we want to focus in particular on the special features that characterize VR as a medium of simulation and the special moral concerns the medium itself may thus present. We turn to an exploration of that medium now.

Virtual reality as a simulation medium: virtually real experiences

Although VR is still a relatively new technology, there is already a considerable amount of data on it that clearly suggests a couple of ways in which VR represents a genuinely new and extraordinary simulative medium. In particular, VR technology can produce experiences that are a great deal more immersive than any other medium to date. By immersiveness, we mean the sense VR users report of being actually transported by the VR equipment out of the space they are genuinely inhabiting in the real world and into the virtual space generated by the equipment, where it then feels like they are really interacting with objects that are in fact only virtual.⁶ When the degree of immersion is strong enough, there is also evidence that VR users sometimes experience a strong sense of embodiment in the VR environment as

⁵ The literature on the possible social and psychological effects of media like films and video games is vast, and we do not propose to give a thorough survey here. The following sources are, however, representative of the sort of work we have in mind: Krahé and Möller (2010) finds some increase in violent behavior among adolescents who engage in violent gameplay while Fischer et al. (2009) finds an even larger increase in violent behavior from those who play games that allow you to customize your avatar. Valkenburg and Peter (2013) offers a general model that aims to explain how cognitive, emotional, and “excitatory” features of games can help to explain why media, especially violent media, affect people differently. Two meta-analytic studies reach opposed conclusions about the correlation between violent media and aggression: Anderson et al. (2010) suggests such a correlation, while Savage and Yancey (2008) resists that conclusion.

⁶ Sanchez-Vives and Slater (2005). Psychologists call this sense of being actually transported into a virtual space “presence”; we discuss this concept of presence and its relationship to our concepts of perspectival fidelity and context-realism below.

well. Even when VR users inhabit a virtual environment in which they are presented with body parts they control that are very different from their actual body—for instance, tentacles instead of hands—they show a shocking ability to treat the virtual analogs of their own bodies as though they are in fact their actual bodies.⁷ Moreover, mounting evidence shows that VR users not only subjectively feel like their bodies are actually being exposed to the events transpiring in the virtual environments they inhabit, but their involuntary biological reactions also suggest that their bodies are often treating those virtual experiences as if they were real⁸; when using VR, we often feel that what happens to our virtual analogue happens to *us*.

We will call such experiences “virtually real experiences.” Such experiences are treated by subjects *as if* they were real experiences (via some combination of behavioral, physiological, neurological, or cognitive similarities between virtual and real experiences). Virtually real experiences are important because of how they affect the nature of VR experience from the point of view of the experiencing subject (Sanchez-Vives and Slater 2005). “When users [have a virtually real] experience..., they feel that the technology has become part of their bodies and that they are experiencing the virtual world in which they are immersed. Moreover, when they feel present in VR, they react emotionally and bodily (at least to some extent), as if the virtual world exists physically” (Parsons et al. 2017).⁹ These virtually real experiences, as well as their characteristic features, are matters of degree. It is possible for some experiences to be more virtually real than others.

We claim that the key features of VR simulations that are essential to the production of virtually real experiences are what we call “perspectival fidelity” and “context-realism.” Although much of what we say about virtually real experiences is empirically contentious, we believe that the available evidence speaks in favor of our position. In the following subsection, we draw upon empirical work to sketch out a theory of perspectival fidelity and context-realism before discussing the unique ethical challenges that virtually real experiences can cause.

⁷ Won et al. (2015). Though subjects may *feel* like their virtual avatars belong to them, we want to distinguish a subject’s perception that she has a tentacle from the perception she would have if she actually had a tentacle. VR may provide subjects with the former but not the latter perception.

⁸ Sanchez-Vives and Slater (2005) and Fox et al. (2012).

⁹ We would like to distinguish experiences of presence from virtually real experiences. Although all virtually real experiences require a subject to experience presence, many experiences of presence will lack the context-realism and perspectival fidelity that we argue are distinctive of virtually real experiences. We say more on this distinction later.

Perspectival fidelity

Perspectival fidelity, as it relates to virtually real experience, refers to the degree to which a representation accurately depicts the subjective point of view of a neurotypical human being¹⁰ and is only nominally related to the *content* of the virtual experience. We emphasize that this property is a matter of degree. We argue that representations that are highly faithful to human perspectives are more likely to generate virtually real experiences than representations that are less faithful.

Perspectival fidelity includes elements pertaining to the relative height of the virtual subject, the depth-of-field visible in the virtual landscape, representations of neurotypical color vision and audition, the absence of non-diegetic sound in the virtual environment, and so on. A god’s-eye-view floating above a VR environment is less perspectival faithful than a grounded view of the same environment. A representation with a non-diegetic symphonic soundtrack is less perspectival faithful than one with only diegetic sound effects.¹¹ Perspectival fidelity is also affected by whether the virtual experience is represented in the first or third person. For example, we conjecture that VR recreations of Philippa Foot’s trolley problem (1978) or Henry Shue’s ticking time-bomb scenario (1978) will be more likely, all things being equal, to be experienced as virtually real if they are constructed from the first-person perspective instead of the third person.

Hardware elements also appear to play a role in perspectival fidelity; the frame and refresh rates of a display can enhance or detract from the perspectival fidelity of an experience depending on whether they impact a subject’s (conscious or subdoxastic) point of view (Sanchez-Vives and Slater 2005). Additionally, the weight of a head-mounted display, to the degree that it becomes part of a subject’s (conscious or subdoxastic) experience, counts against the perspectival fidelity of that experience (Ramirez 2017).

¹⁰ We use the term “neurotypical” here as a descriptive statistical term to denote the range of sensory capacities available to the average adult human being. We embrace what some have referred to as ‘neurodiversity’ movements (Herrera and Perry 2013) and do not intend to use the term neurotypical normatively. Deafness is, by all accounts, not neurotypical though arguably it is not a disability or disease for those in the deaf community. An experience that lacks auditory inputs, however, is less perspectival faithful than one that includes such inputs. Perspectival fidelity will relativize to the typical phenomenology of the subject population (e.g., perspectival fidelity for gorillas will look differently than for neurotypical humans and perspectival fidelity for the deaf will vary in many respects from that of hearing persons).

¹¹ For similar reasons, such a soundtrack would diminish the degree of context-realism of the representation.

Context-realism

The second element we believe composes virtually real experiences is *context-realism*. Like perspectival fidelity, context-realism is a matter of degree; some experiences can be more context-real than others. Where perspectival fidelity refers to features relevant to the structure of experience, context-realism refers to features relevant to the *content* of an experience. Such features may include aspects relating to the plausibility, from the subject's point-of-view, of a virtual world's environment. The more a virtual world's environment is bound by the same physical and psychological principles that a subject believes grounds their own world, and the more these rules cohere with a user's lived experience, the greater the context-realism of that environment and the more likely it is for such an environment to produce virtually real experiences for subjects. Interestingly, context-realism does not appear to depend heavily on photo-realism when it comes to producing virtually real experiences. In other words, the degree of graphical realism of a representation does not appear to dramatically affect or determine a subject's beliefs about the plausibility of an environment (Sanchez-Vives and Slater 2005; Slater et al. 2006).

All sorts of features can make a simulative environment more or less context-realistic. For example, a world in which a subject is able to fly unassisted is less context-real than one where this is not possible. A game where flashing power-ups and bonuses fly out of a defeated opponent's corpse is less context-real than one that handles death more realistically. An environment designed with graphical overlays, voice-overs, or other forms of meta-content will be less context-real (and simultaneously less perspectivally faithful) than an environment that lacks these elements.¹² Additionally, we conjecture that environments set in distant (or alternative) pasts or futures will be less context-real than those set contemporaneously relative to the subject; they will be more likely to be treated as fantasy or game worlds rather than real spaces. Kinesthetic elements can also, we believe, add or detract from a simulation's context-realism. For example, a simulation of nocking an arrow onto a bow that requires the subject to physically move her arms in a nocking motion is more context-real than one in which the same action is carried out with a keyboard or mouse.

Although context-realism is a measure of the physical and psychological rules of a virtual environment, such a measure contains some elements that are subjective and some which will be objective (owing to shared features and limitations

¹² As augmented reality devices become more widespread and such meta-content becomes a standard component of lived experience, simulations that include this sort of meta-content may thereby become *more* context-real.

of human embodiment). For example, an important component of context-realism is determined by a subject's beliefs about what is possible in our world. A supernatural virtual environment, we conjecture, will be more likely to generate virtually real experiences for subjects who believe that supernatural features are real in the actual world than in subjects who do not share these beliefs. A virtual environment set in a haunted house is more likely to generate virtually real fear (as opposed to other forms of fear) for someone who believes ghosts are an actual part of the fabric of their universe.

Context-realism is relative in another important sense as well. The features of an experience that render it context-real for some purposes may be different than those that render it context-real for other purposes. For example, the features of a situation that would produce a virtually real experience of a moral dilemma may differ from the features necessary to produce an erotic virtually real experience or an aesthetic virtually real experience.¹³ However, we also believe that some elements of context-realism will be much more widely shared, if not universal, owing to the nature of human experience and human embodiment. One plausible candidate for such a feature is psychological. Virtual agents who are moderately reasons-responsive, in Fischer and Ravizza's (1998) sense, are likely to add to the context-realism of a scenario, whereas virtual agents who are not will likely detract from it. For example, when Mel Slater and his colleagues (2006) attempted to create a virtual replication of Milgram's (1963) obedience studies, their virtual 'learner' was a blocky, photo-unrealistic woman sitting on a chair (see Fig. 1).

However, the learner was programmed to respond in psychologically plausible ways to receiving shocks which, coupled with the high perspectival fidelity of the virtual scenario, induced virtually real experiences in subjects:

When the Learner failed to answer at the 28th and 29th questions, one participant repeatedly called out to her 'Hello? Hello? ...' in a concerned manner, then turned to the experimenter, and seemingly worried said: 'She's not answering ...' In the debriefing interviews many said that they were surprised by their own responses, and all said that it had produced negative feelings. (Slater et al. 2006)

We will say more about the ethical concerns that virtual environments like these should raise for Institutional Review

¹³ This is an empirical conjecture on our part. As we noted above, we do not wish to produce a view on the necessary and sufficient conditions for context-realism or perspectival fidelity. That is a task better suited to psychologists and neuroscientists. What we do wish to do is to mark out the concept of virtually real experience and its connection to VR experience in order to generate what we believe is a novel and underappreciated ethical concern about such experiences.

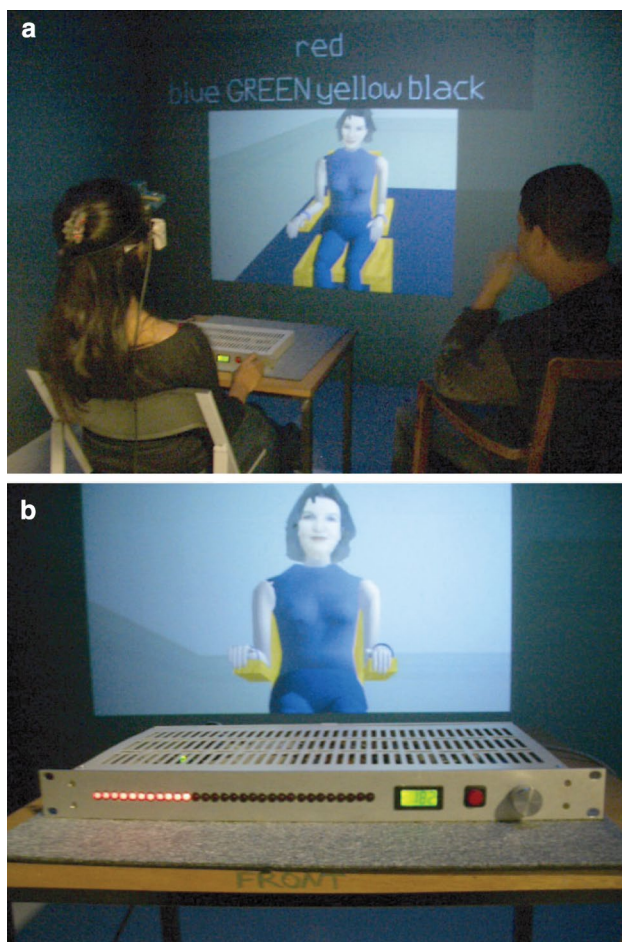


Fig. 1 The virtual ‘learner’ used by Slater et al. (2006)

Boards in a later section. For now, we note that most virtual agents do not behave, even remotely, in moderately reasons-responsive ways. This may be one reason why violent content (virtual and PC-based) is typically not traumatizing for subjects. Such environments lack context-realism to a degree that blocks the production of virtually real experiences and prevents subjects from engaging with the content of these experiences as if they were real.

Virtually real experiences are experiences that are treated *as if* they were real by the experiencing subjects. In part this claim is grounded on physiological and behavioral measures. For example, in the example drawn from Slater et al. (2006) above, subjects not only behaved *as if* they were agitated by the experience, they also expressed concern verbally for the virtual learner. Beyond behavioral assessments, physiological measures of subject experience are often used to assess the nature of subjective experience (Kivikangas et al. 2011). Using such measures, experimenters observe similarities between physiological markers (especially EEG, skin conductance, and EMG) subjects display during both real experiences and what we are calling virtually real experiences

(McLay et al. 2011; Felnhofera et al. 2015; Higuera-Trujillo et al. 2017). In this sense, although participants may offer post-hoc reports that they did not believe that their VR experiences are real, we argue that their experience-in-the-moment is being treated *as if* it were real.¹⁴ Self-reported experiences of presence do not always correspond to physiological responses indicating presence (Felnhofera et al. 2015). Perhaps the strongest evidence for the existence of virtually real experience lies in the effective use of VR Exposure Therapy (VRET), especially in comparison to traditional imagination-based therapy (McLay et al. 2011; Parsons and Rizzo 2008). We claim that such experiences are more likely to be generated by VR environments because these environments can demonstrate levels of perspectival fidelity unavailable in other media. When coupled with highly context-real worlds, such experiences are likely to be virtually real. The interactions between perspectival fidelity, context-realism, virtually real experiences, and psychological effects on users will be central to the moral lessons we wish to draw about VR.

Virtually real experiences and presence

One way in which psychologists have described the immersiveness of media like VR is via the concept of what they call “presence”, which was defined in one study as “the sense of being in a [virtual environment] rather than the place in which the participant’s body is actually located.”¹⁵ We want to briefly distinguish the psychologists’ concept of presence from the concept of virtually real experience.

Conceptions of presence are contentious. Indeed, on some accounts, an experience high in “presence” and a “virtually real” experience are roughly synonymous:

[P]resence in a [virtual environment] is inherently a function of the user’s psychology, representing the extent to which an individual experiences the virtual setting as the one in which they are consciously present. On the other hand, immersion can be regarded as a quality of the system’s technology, an objective

¹⁴ We are interested primarily in how subjects experience VR simulations in-the-moment, as it were. Our analysis, therefore, focuses on virtually real experiences and describes those as experiences that are treated *as if* they were real in the moment they are being experienced. A treats VR experience b as if it were real if A, either behaviorally, physiologically, neurologically, or psychologically reacts to b in a similar way as they would react to a real-life experience of b. It is entirely possible that subjects may re-frame these experiences after the fact (“it wasn’t real anyway”). We believe that moral issues can arise with respect to how subjects process their experiences after-the-fact, though these are outside the scope of this article. We thank an anonymous reviewer for helping us clarify this concern.

¹⁵ Sanchez-Vives and Slater (2005, p. 333).

measure of the extent to which the system presents a vivid virtual environment while shutting out physical reality. (Cummings and Bailenson 2016)

If, however, presence is understood as involved “in the process of perspective taking, the mental simulation of a situation by placing oneself in the shoes of another via imagination” (Ahn et al. 2016), then presence and virtually real experiences come apart. Elsewhere, we have argued that subjects may feel present in a simulation while being fundamentally mistaken about the content of that simulation (Ramirez 2017). For example, a subject may believe they are experiencing what it is like to be a cow in a slaughterhouse (Ahn et al. 2016), a highly present experience, while being fundamentally mistaken about the nature of that experience (e.g., they are not genuinely experiencing what it is like to be a cow). Therefore, although in Ahn et al.’s (2016) understanding of the role of presence, the experience of presence is a necessary feature of what we are calling virtually real experiences (i.e., all virtually real experiences are experiences involving presence), nevertheless many experiences of presence are not experienced as virtually real by subjects. Although subjects may report feeling like they are genuinely in the VR environment, they can be fundamentally mistaken about the perspective that they are occupying.¹⁶

Virtual reality and other simulative media compared: a ride on the bus

VR’s capacity to generate virtually real experiences distinguishes it from every other simulation medium developed to date, and it is our view that under some circumstances, this feature affects the experience a simulation’s subjects might have such that simulations that are morally unproblematic in other media become problematic in VR. To capture the salient circumstances that make the moral difference, we propose to compare how various simulative media might simulate a specific set of events, and how those simulations might be experienced differently by the simulations’ subjects. These events are provided by a well-known series of philosophical thought experiments, Joel Feinberg’s “Ride on the Bus” (Feinberg 1985, pp. 10–13). Feinberg originally created these thought experiments to prompt his readers to reflect on their responses to a variety of offensive but (purportedly) harmless experiences, the question being, in which cases would we be morally justified in demanding

legal protection from these offenses, even at the cost of limiting other people’s liberties? Feinberg sets the experiment up thus:

In each story the reader should think of himself as a passenger on a normally crowded public bus on his way to work or to some important appointment in circumstances such that if he is forced to leave the bus prematurely, he will not only have to pay another fare to get where he is going, but he will probably be late, to his own disadvantage. If he is not exactly a captive on the bus, then, he would nevertheless be greatly inconvenienced if he had to leave the bus before it reached his destination. In each story, another passenger, or group of passengers, gets on the bus, and proceeds to cause, by their characteristics or their conduct, great offense to *you*.¹⁷

Feinberg goes on to enumerate several different sorts of offenses you might then find yourself subjected to on the bus. This is not the place to quote every offensive story in full, but here is a summary of Feinberg’s typology of offense:

- Affronts to the senses: Violently clashing clothing, loud and obnoxious sounds (e.g. nails on a blackboard)
- Disgust: People engage in crude behavior like farting and belching; they eat revolting things; they then vomit them up, and finally eat each others’ vomit and feces
- Shocks to our Sensibilities: People defile a human corpse or important symbols, or wear clothing with shocking words or images
- Shame and anxiety: People are nude and engage in public sex (including all sorts of variations involving all sorts of partners, human and animal)
- Boredom: You are forced to listen to endless, insipid conversation, or are chatted up by an annoying person who won’t leave you alone
- Fear: Other passengers pretend to use toy weapons to harm others, or possess clothing or signs with messages or symbols that arguably threaten harm or signal hatred toward various groups

All of these events are things we could imagine simulating in a variety of media. How might these simulations be experienced in different forms, and what sorts of ethical issues arise regarding the appropriateness of creating such simulations?

We consider four different forms in which the ride on the bus might be experienced: a traditional classroom presentation of the material in print and lecture, a conventional film,

¹⁶ For similar reasons, we have doubts about the ability of VR environments to allow for any form of “in-their-shoes” empathic perspective taking (Goldi 2011; Ramirez 2017), though we do not deny that subjects in such environments feel a high degree of presence in them.

¹⁷ Feinberg (1985, p. 10).

a conventional interactive video game, and a VR simulation. The first of these, the classroom presentation, we do not consider to be a simulation, properly speaking; while reading Feinberg's text of the thought experiment may be intended to provoke the audience to imagine what the described experiences might be like, there is no meaningful sense in which the lecturer would be trying to present the audience with anything that would *feel like* the described experiences themselves. Several features characterize this sort of presentation and differentiate it from the simulations that follow. First, the classroom presentation offers the audience only thin background content in describing the events on the bus. The verbal description of the scenario explicitly describes only the foregrounded material, requiring the audience to fill in a lot of the details imaginatively: What do the bus, the bus driver, and the passengers look like? What is the temperature on the bus? How widely are the bus seats spaced? What side of the bus should you imagine yourself sitting on? Second, this presentation of the events on the bus is relatively escapable. Although it confronts the audience with topics and descriptions they might not choose to conjure up for themselves, once these things have been brought up, it is not difficult for the audience to control for themselves the intensity with which they continue to confront themselves with offensive or unpleasant content. They can presumably stop listening or stop imagining the offensive elements whenever they like. Finally, the presentation completely lacks perspectival fidelity: there is no sense in which the audience is being given an experience that accurately provides them the perspective they would have were they actually on the bus, and there's no question of their separation from the events being described. They are not genuinely "in" the described environment in any meaningful way. The whole experience boils down to words spoken and images generated in the audience's head.

The second medium, film, we imagine can constitute a genuine simulation of the ride on the bus¹⁸; film is a medium that can attempt to give viewers some experience of what it would actually be like to inhabit a particular place and time. Note, however, that some of the events described on the bus would not be merely simulations in a film, but would instead be unsimulated *instances* of the offenses under discussion; the sound of nails on a blackboard would still be the sound of nails on a blackboard, for instance, and insipid conversation one is forced to listen to in a film is still insipid

¹⁸ Note that our analysis of the potential harms of a film would be different if the film were not a simulation of fictional events but a documentary of actual events; the ethics of filming and viewing a simulation of someone being stalked and killed, for instance, are, we assume, different from the ethics of filming and viewing an actual murder. In what follows we will be focusing our attention entirely on simulations.

conversation, not just a simulation of it.¹⁹ If such real-life experiences are inherently offensive, cinematic presentation of them is presumably offensive as well. But beyond this, cinematic simulation displays the following features: First, it provides thicker background content than the classroom presentation. The director of a film about the bus would have to decide, for instance, what sort of bus to use, who the bus driver would be, and so on, entailing that much of the work of imagining background detail would be taken out of the hands of the audience. Second, the film is decidedly not interactive; it will play exactly the same way every time we show it, and the audience observes the film passively. Third, the film's presentation (barring a *Clockwork Orange* sort of scenario) is still relatively escapable; it may be harder to hide from the events simulated in a film than it is from a verbal classroom description of those events, but one could still turn away, close one's eyes, cover one's ears, etc. Finally, while the film might offer a degree of perspectival fidelity—for instance, it might be shot from the point of view of a rider on the bus whose perspective you might inhabit—nevertheless, film audiences will generally feel quite separate from the actions depicted on screen. There will always be a frame to the film, and my experience will always be of a screen-bound "film-in-the-world."

A conventional computer simulation of the ride on the bus could share many of the features of a film, with some new features added. Like a film, a video game could supply thick background content, supplying all sorts of details about the environment on the bus. Like a film, it would still be a relatively escapable experience; one could always back away from the computer, and we would still experience it as contained within the frame of a monitor screen. On the other hand, a typical computer game will be interactive to some degree; it will allow the player the ability to make some choices or exert some control (e.g. over perspective and movement, perhaps allowing interaction with simulated objects) in the simulated environment. And perhaps because of this interactivity, the environment might be more context-real; the player will typically have a stronger feeling that she is "in" the simulated environment than would be the case in the non-interactive, passive experience of the film. Such experiences are, however, unlikely to be experienced as if they were real. In part this is because such experiences are

¹⁹ Strictly speaking, the sounds of a nails on a blackboard or of insipid conversation would be recreations or instances of an *aspect* of an experience but would not by themselves rise to the level of a recreation or instance of the *whole experience*. This is because of the different situational factors (subdoxastic elements of experience) that would be missing from the film version of the experience relative to the first-personal experience of being on a bus. For example, while a threatening gesture aimed at the camera may be visually similar to the same threatening gesture aimed at you in reality, your experience of the two gestures is likely to be qualitatively different.

still relatively escapable. Additionally, such experiences usually include many cues that often make games less context-real than they could be. These are likely to prevent a player's sense of presence from being strong. Still, more than in the case of film, a computer simulation may demonstrate a relatively high degree of context-realism. The rules of the simulated environment might model in a relatively faithful way the rules of our own world; simulated objects might behave as real-world physical objects do, for instance, and such realism could increase the degree of presence subjects feel.

Turning finally to VR simulations of the ride on the bus, we find VR shares some features with film and computer environments. It shares thick background content with both media, and it shares computer simulations' potential for interactivity. Indeed, some VR systems that use handheld controls can enhance interactivity by enabling manipulation of virtual objects in real space with real physical movements; for instance, one can use one's actual hands to physically draw and aim a virtual bow one is holding in one's virtual hands, and even feel the controller physically shake to mimic the tension in the bowstring. And all of this can contribute to the high degree of perspectival fidelity and (perhaps) context-realism we have attributed to VR simulations; first-person VR simulations of the ride on the bus could locate you on a virtual bus seat, surrounded by a virtual bus, a virtual driver, and virtual fellow passengers on all sides, all experienced in surround sound and presenting neurotypical field of vision and depth perception. Furthermore, while films and computer simulations can aspire to replace a subject's point-of-view with the point of view of a subject in a virtual environment, these media are always screen-bound (they are always experienced as being contained by a screen within the larger field of view of the user). VR systems are not screen-bound in this way, and so the perspective shifting they introduce is more complete. VR systems are designed to "completely occlude any visual contact with the outside world and replace it with computer-generated images, which are dynamically adapted to any viewing position by means of head-tracking" (Parsons et al. 2017). The fact that screens are components of all PC and filmic experiences will, we conjecture, always reduce the perspectival fidelity of these experiences relative to an otherwise identical VR depiction. We believe that this feature of VR experiences lends itself to the creation of virtually real experiences to a degree missing from other forms of media, and this feature will be the source of several of our ethical concerns. Moreover, the very fact that VR equipment is worn mounted on one's head and completely fills the user's field of vision makes this simulative medium harder to escape than any film or computer game confined to a stand-alone, stationary screen. There's no turning away from a VR simulation, since one's point of view simply tracks the motion of one's head to reveal a new part of the virtual landscape. If part of what Feinberg

wanted us to imagine for ourselves when he described the ride on the bus is that we are *stuck* on the bus such that it's hard to get away from the ride's many offenses, VR can simulate that inescapability more realistically than any other medium to date.

Taking all these experiential differences between simulative media into account, what moral judgments should we make about simulating the various offenses described in Feinberg's bus ride? Our answers here are provisional, and readers are invited to disagree with our particular moral pronouncements. We are more interested, in this section, to argue that such moral judgments are likely to track features of a situation that will *differ* across modes of presentation (imagination, film, gaming, VR) even if the *content* of the simulation remains the same.²⁰

We will suggest that the offenses in Feinberg's thought experiment fall into three categories. First, some of these offenses seem to us to be of a sort such that we should not object to their description or simulation in any format. In this group we would place, for instance, people wearing violently clashing clothing, farting, and belching; these just don't seem to be the sort of experiences that, to us, would be likely to traumatize their subjects in any serious way, no matter the medium.²¹ A second category of offense seems to be problematic purely because of specific offensive content, like bigoted messages, and in these cases we suggest the morally problematic character of any simulation would be due not to the nature of the medium, but purely to the offensive content itself. We would place many of Feinberg's "shocks to sensibilities" in this category: he imagines, for instance, a t-shirt with a picture of Christ on the cross with a caption underneath reading "Hang in there, baby!", or a "banner with an offensive caricature of a woman and the message 'Keep the bitches barefoot and pregnant'" (Feinberg 1985, pp. 11, 13). Similarly, aspects of an experience which above we described as not so much simulated as

²⁰ In assessing these offenses, we set aside the very real issues that others have raised with encouraging subjects to themselves engage in unethical behavior, concerns sometimes discussed in terms of the "Gamer's Dilemma." Although we agree that simulations which encourage subjects to rape, torture, or kill virtual persons raise important ethical issues (especially in terms of long-term effects on individual and societal norms), we sidestep this concern here to focus on the nature of the subjective trauma that may be experienced by the subject of the experience herself. Thanks go to an anonymous reviewer for asking us to clarify this concern.

²¹ As VR technology develops, it is possible that some things that are only simulatable now might become reproducible, and smells seem a likely candidate. If VR simulations someday include elements like reproducing the odor of flatulence, and if we arrive at a consensus that being exposed to reproduced flatulence is so unpleasant that people would reasonably want to be protected from the experience, we would have to consider moving such offenses into a different category.

reproduced—the sound of nails on a blackboard, inescapable and insipid conversation, extremely loud music—would be potentially problematic (or not) in whatever medium they are reproduced. We do not propose in this paper to discuss which content is morally unacceptable to inflict on subjects and which is not; our readers are welcome to employ their own moral judgments on the matter. Rather, our point is that in these cases, specifically the *medium* of the content's simulation (or perhaps even simply of its description?) seems not to be morally relevant.

In a third category we would place a range of offenses that start being morally objectionable when they are realistically simulated, and become more objectionable as the simulation becomes more inescapable and likely to be experienced as virtually real. To our minds, the paradigm examples of this category of offenses involve sex and violence. Imagine, for instance, being a college professor deciding what sorts of depictions of sex acts she could show her class. No doubt some professors would not be comfortable even describing verbally the panoply of sex acts Feinberg mentions, but as we imagine moving from the contemplation of crude, cartoonish renderings to more and more lifelike and realistic depictions of these acts, adding animation, a soundtrack of realistic noises, and even ultimately the possibility of interactivity, we expect fewer and fewer professors could bring themselves to share such simulations with their students. We harbor a similar intuition where the simulation of violence is concerned. These cases are similar to those in the second category in that one's moral assessment of the simulated content is going to largely determine one's assessment of the simulation; someone who does not morally object to pornographic content generally, for instance, might not see a moral problem with any of these simulations. But for someone who does have such an objection, we expect that objection will become more and more strenuous as the simulation becomes more realistic, more inescapable, and more likely to be experienced as virtually real by students. We would expect such a person to think, for instance, that a perspectively faithful and context-real VR simulation of sex acts would be more morally problematic than a pornographic film treating the same material, and that both would be more problematic than a written description.

Two more cases from Feinberg's examples make for especially apt instances in which the perspectival fidelity of VR is likely to increase the likelihood of generating virtually real trauma: eating disgusting things like vomit and feces, and

the violent defiling of a corpse.²² Take the latter case, which Feinberg describes thus:

A group of mourners carrying a coffin enter the bus and share a seating compartment with you. Although they are dressed in black their demeanor is by no means funereal. In fact they seem more angry than sorrowful, and refer to the deceased as 'the old bastard' and 'the bloody corpse.' At one point they rip open the coffin with hammers and proceed to smash the corpse's face with a series of hard hammer blows. (Feinberg 1985, p. 11)

To be sure, it could be disturbing to witness such an event simulated in a film or computer simulation, but both remain screen-bound experiences (the screen would form only one part of the subject's overall field of view). As a result, the low perspectival fidelity of such experiences will *lessen* their likelihood of inducing subjective trauma; in particular, extant research suggests that HMD-mediated experiences are far more likely to produce physiological and behavioral responses matching real experiences than those that are merely imagined (McLay et al. 2011; Riva et al. 2007; Felnhofera et al. 2015).²³ Although film, PC, and VR simulations can all present us with context-real depictions of Feinberg's scenarios, the ability of VR simulations to offer unprecedented degrees of perspectival fidelity make it more likely that these simulations will produce virtually real experiences of those simulations. If it would be traumatic to experience either of these scenarios on a real bus, we argue that the virtually real experiences of these scenarios that VR is capable of producing are likely to also induce trauma.

Our own intuitions about which cases may be more likely to produce virtually real trauma notwithstanding, we argue that virtual experiences high in context-realism and perspectival fidelity are more likely to produce virtually real experiences than their merely imagined counterparts. Many of the scenarios on Feinberg's bus include elements like incest, bestiality, and so on which psychologists have argued are intrinsically disgusting and perceived as immoral across cultures (Haidt et al. 1993). Our arguments, however, rely only on the assumption that at least *some* of the scenarios on the bus are likely to be traumatic if experienced in the real world

²² Our point becomes even stronger if we assume olfactory elements can be introduced to these simulations. However, even if a reader thinks these particular cases are still not morally problematic when experienced as virtually real, so long as they can imagine a scenario in which an experience becomes morally problematic when it becomes virtually real, the argument progresses.

²³ Also, for a modern audience watching a film, part of the phenomenology of viewing a film involves the consciousness that the events one is watching on-screen were filmed at some point in the past, and so are not genuinely present. This is not true of traditional and VR computer simulations.

and therefore that they are also capable of inducing trauma if experienced as virtually real in VR simulations, especially those high in both perspectival fidelity and context-realism.

The equivalence principle

As we have seen, some of the potential moral issues raised by a VR simulation of the Ride on the Bus are shared by its film and text-based cousins. Nevertheless, though our imaginations may be vivid, they are constitutively incapable of the forms of perspective-taking that thought experiments like Feinberg's require (Ramirez 2017). For this reason, text-based scenarios are unlikely to produce virtually real experiences.²⁴ Additionally, though film and computer experiences may produce stronger reactions in subjects than text-based scenarios, the passive and screen-bound nature of these media make it less likely that they will generate virtually real experiences. Philosophers have long wrestled with the limits of the imagination, and here we see ourselves contributing to this tradition.²⁵ In this section, we highlight what we believe are unique ethical issues that arise from the tendency of some VR environments to generate virtually real experiences in users and propose a heuristic tool to help assess the risks involved in VR simulations.

For example, Institutional Review Boards (IRBs), whose function is to assess whether the use of human subjects in research is appropriate, are extremely wary of attempts to replicate Milgram's (1963) obedience experiments. In part, these concerns are grounded on the fact that subjects of these experiments were deceived about the nature of the study; additionally, subjects were exposed to an extremely stressful, potentially traumatic environment to which they could not consent prior to the study and in which they felt some pressure to continue (Schlenker and Forsyth 1977). Some researchers are using VR to conduct experiments which they

hope will generate ecologically valid data, and at times they are taking advantage of virtual environments to sidestep ethical concerns that have confronted traditional lab-based experimenters. VR environments are currently not thought to be as dangerous or potentially traumatizing as analogous real world environments. Many researchers see VR as capable of allowing "all social and psychological research where, for ethical or safety reasons, it is not possible to immerse experimental participants into the actual phenomena to be studied" (Slater et al. 2006).²⁶ Some go so far as to claim that "virtual reality allows us to conduct experiments that would be ethically unacceptable to execute in non-virtual environments" (Skulmowski et al. 2014). We strongly disagree with this way of understanding the promise and perils of VR technology.

There are reasons beyond those we are focusing on for worrying about the effects of VR on subjects. A number of researchers have proposed several ethical concerns that they believe are unique to VR media. For example, some have argued that the feeling of presence allowed by VR media may cause permanent psychological or biological changes in subjects, including the neural mechanisms responsible for proper embodiment (Madary and Metzinger 2016). Others have noted that the particular *ways* in which users are embodied in VR can have long term effects on behavior; for example, users embodied in aged avatars were more likely to save for retirement (Rosenberg et al. 2013), and it is not hard to imagine less beneficent effects other VR experiences might produce. However, while we believe that such ethical concerns are warranted, we also believe, in addition, that the nature of virtually real trauma has gone unnoticed and deserves greater scrutiny.

Given our claims about context-real and perspectivally faithful VR environments and their ability to generate virtually real experiences, we argue that the same level of scrutiny should be applied to all faithful VR replications as is applied to lab-based experiments like Milgram's obedience studies. Indeed, we believe that virtually real experiences, once appreciated, should lead IRBs and designers of VR simulations to apply a heuristic device that we call TEP when designing virtual worlds:

²⁴ Although the products of imagination are almost always incapable of the sort of perspective-taking that produces virtually real experiences, they are capable of triggering trauma in some subjects. This is a significant concern and we do not wish to downplay it. Such scenarios' ability to induce trauma appears not to depend on their medium (text, film, VR), and so we do not focus on it in this paper. It should, however, remain a real concern for those who wish to expose naive subjects to potentially traumatic scenarios in any form.

²⁵ For example, our concern, stated very generally, is about a form of "imaginative resistance": "imaginative resistance occurs when an otherwise competent imaginer finds it difficult to engage in some sort of prompted imaginative activity" (Szabó and Liao 2016, p. 405). In our case, however, we argue that the problem runs deeper than finding it "difficult" to imagine the scenarios of these thought experiments. Specifically, we believe that features of first-personal perspectives themselves can make it all but impossible to carry out these thought experiments via the imagination (Goldie 2011; Ramirez 2017). We thank an anonymous reviewer for this clarification.

²⁶ In fairness, Slater et al. (2006, p. 7) appear to appreciate this concern: "[t]he actual conditions of Milgram's experiments can, of course, never be exactly replicated in virtual reality since the participants will always know that the situation is unreal—and if eventually virtual reality became so indistinguishable from reality that the participants could not readily discriminate between the two, then the ethics issue would arise again." However, they fail to appreciate that virtually real experiences are dimensional and may be generated even without photorealistic environments. Their own research provides evidence for this claim.

TEP: If it would be wrong to subject a person to an experience then it would be wrong to subject a person to a virtually real analogue of that experience. As a simulation's likelihood of inducing virtually-real experiences in its subject increases, so too should the justification for the simulation's use.

TEP is not a new ethical principle. All scrupulous researchers already show concern about the relative harms that their experiments may cause their subjects, and seek to minimize those harms. Additionally, game designers (if only for liability reasons) will want to avoid non-consensually harming their users. TEP helps, we believe, to highlight the important way in which perspectively faithful and context-real simulations impact user experience in potentially harmful ways. It is, in part, meant to induce caution and care in the creation and use of VR environments that can generate virtually real experiences. Like other researchers who are concerned with VR's potential to cause intentional and unintentional harms, TEP pushes back against those who believe that because an environment is virtual it cannot be harmful in the same ways laboratory environments can be.

Our appeal to TEP as a heuristic guiding principle should not be read as a call for ending VR research in morally fraught areas or the creation of games with morally complex situations. Indeed, we believe that we must continue to explore how human experiences are shaped in VR not only in terms of subjective trauma but with respect to long term cognitive, affective, or behavioral consequences (Madary and Metzinger 2016). TEP, as a heuristic guiding principle, is positioned to help researchers and designers generate appropriate caution toward the creation and implementation of VR-centered environments. If an experimental protocol would raise ethical issues if performed in a laboratory setting, then TEP, which reflects pre-existing commitments to avoid unnecessary harm, demands that we raise the same ethical issues for a VR version of the same protocol that is high in perspectival fidelity and context-realism.

One advantage of a heuristic like TEP is that it is substantively neutral between competing normative ethical theories. In other words, TEP is a 'thin' principle (Williams 1985). Whatever your preferred normative or moral framework, TEP relies only on the intuition that one ought to treat like cases alike. It is fueled by the empirical claim that some VR environments generate virtually real experiences and that these experiences produce reactions (behavioral, autonomic, and cognitive) similar enough to their real-world analogues that we should consider treating them similarly. The need for justification to engage in such research would therefore grow in proportion to the supposed risk of harm that such VR simulations would generate.

Institutional Review Boards, as mentioned above, are tasked with assessing the costs and benefits of research

before granting (or denying) approval for the research to continue. Research involving human subjects falls under even more scrutiny wherein harms must be minimized to subjects. As we argued above, we believe that IRBs should make use of TEP in order to aid their work with identifying and adjudicate potential harms involving the use of human subjects research that makes use of VR environments.

Any virtual environment capable of generating virtually real experiences should fall under the same scrutiny as an experiment carried out in a more traditional laboratory setting. We believe that it is especially important for IRBs to institute a heuristic guiding principle like TEP sooner rather than later as VR research becomes more common. Intentionally or not, the design of Slater et al.'s (2006) replication of Milgram's experimental protocols was able to elicit virtually real experiences in subjects. As the software and expertise for building context-real and perspectively faithful virtual environments becomes cheaper and more widely available, the risk of accidental trauma resulting from human subject research is likely to grow. TEP is structured to minimize this risk.

A second advantage of TEP, we believe, is that its scope is limited. TEP would not apply to all VR environments. Given the current state of the technology and the dominance of gaming applications in VR, it is likely that TEP would apply to only a fraction of current VR environments. TEP is concerned only with virtual environments that, intentionally or unintentionally, generate virtually real experiences. VR environments where subjects take on the role of an unstoppable killing machine, while potentially objectionable for content reasons, fall far outside the scope of TEP. Such scenarios are simply too context-unreal.²⁷

Finally, a third advantage of formally articulating and defending a heuristic like TEP is to make clear how our pre-existing commitment to avoid unnecessary or accidental harm actually plays out with respect to VR technologies. A heuristic like TEP thus brings greater awareness to the role of context-realism and perspectival fidelity in VR environments. Moreover, these concepts can be of general application throughout the world of VR development, which

²⁷ While it would surely be wrong to amputate a subject's healthy limbs in real life even if the subject consented, surely it is not wrong (at present) to simulate lopping off limbs in VR. On our view, this is true only given the limitations of existing VR technology. If in the future companies produce VR bodysuits with the capacity to, for instance, inflict high levels of pain on their wearers, we might well decide it is no longer morally acceptable to simulate experiences that cause extreme pain in VR. As the levels of context-realism and perspectival fidelity that technology permits increases, we will need to recalibrate our intuitions about what is and is not acceptable to simulate. We thank an anonymous reviewer for pointing out the need for clarification on this point.

suggests a final consequence of our theory that we explore in the following section.

TEP and VR media

TEP's scope is currently limited, in part, by the nascent state of the technology. As the technology grows in popularity and market share, we believe that the need for thoughtful application of TEP will grow. This is especially true regarding VR media of various kinds.

Although VR gaming is likely to lack context-realism (as a result of a game's setting, physical rules, limited artificial intelligence, or its game logic), game designers can utilize context-realism and perspectival fidelity to more carefully shape a player's affective experiences and avoid accidental player trauma. For example, many fans of *Grand Theft Auto 5* (Rockstar North et al. 2013), a (non-VR) game in which players take on the role of three criminals, were surprised, and disturbed, by one of the game's missions. This mission required the player to torture a virtual person in order to extract information about a terror plot:

Dubbed “the most disturbing scene” in the game by the popular gaming site Eurogamer, the episode comes midway through a mission called “By The Book,” which has players oscillate between two of the game's player characters — Trevor and Michael — as they work to hunt down a terrorist at the behest of a fictional version of the FBI. Trevor chooses between using a variety of torture-friendly appliances like a wrench and electrified clamps to persuade the suspect into divulging more information, while Michael uses the intelligence gathered from this enhanced interrogation to hunt down the alleged terrorist and take him out with a sniper rifle. In one particularly ugly moment, the player makes a circular motion with the gamepad's joystick to wrench a tooth out of the suspect's mouth. (Lejacq 2013)

The *Grand Theft Auto* series is infamous for allowing, even encouraging, players to maim and kill civilian bystanders, sex workers, and police officers. Given the intentionally violent context of the game, it might be surprising that players would have such a strong response to the mission above. We believe that perspectival fidelity and context-realism can help explain why reactions like this can occur (intentionally or not). For example, *Grand Theft Auto* designers typically provide the player with a third-person perspective high above the player-character, diminishing the perspectival fidelity of the experience. Additionally, designers overlay unrealistic meta-content (health, cash, weapons and ammunition, a map, etc.) on the player's screen. Further detracting from the context-realism of the environment, bundles of cash occasionally emerge from the bodies of virtual agents the

player has killed, and the player him or herself can take an unnatural amount of damage before (temporarily) dying. All of these features of the *Grand Theft Auto* experience make it less likely to produce virtually real experiences for players in virtue of diminishing the game's degrees of perspectival fidelity and context-realism.

The torture scene described above, however, is unique in that it removes many of the elements detracting from the perspectival fidelity and context-realism of the user experience. The tortured person, unlike all other virtual agents the player interacts with, responds realistically to pain: he audibly screams, winces, and pleads with the player to stop. Meta-content disappears from the screen for this scene, and vital information is delivered diegetically (e.g., the torture victim is hooked up to a heart monitor so players can track how close to death he is); these features both add to the perspectival fidelity and context-realism of the scenario. Instead of the godlike 3rd-person point of view characteristic of the game's normal mode, the player's perspective during the torture sequence shifts so that they are just behind the character they are controlling. The player is given the perspective of someone who is in the room standing just behind their avatar. These features of the mission, we argue, impact the perspectival fidelity and context-realism of the environment enough to produce (intentionally or unintentionally) genuine unease, even trauma, for many players. Understanding how these elements of a game experience can be manipulated to heighten or decrease affective engagement is also useful to avoid accidental player trauma (as may have happened in this case). Happily, there is already evidence that some game designers are alive to the issue.²⁸

Conclusion

For many reasons, the spread and rapid development of VR technology is very exciting. Its potential is vast, including not only the promise of great entertainment but also the promise of therapeutic, artistic, archival, educational, and any number of other sorts of applications. We are as excited about these possibilities as anyone. We also believe, however, that some of the very features that make this new medium so exciting should also make us cautious. The scientific community has only begun to explore the way human psychology interacts with this new technology, but the early evidence suggests that we have a remarkable capacity to imagine ourselves as actually inhabiting some of the virtual environments that digital creators have invented, and this

²⁸ For examples of game developers expressing concern about VR simulation-induced trauma, see Hudson (2016) and Handrahan (2016).

capacity may well bring with it the risk that inhabitants of a virtually real environment can experience very real trauma there. Especially as the results of research about VR begin to accumulate, we urge the community of VR creators, the companies that employ them, and the researchers who are using VR to hesitate long enough to consider the possibility that some virtually real experiences may inflict genuine harm. Careful application of a heuristic principle like TEP would prevent such harms from occurring, and to that end we humbly suggest its use to all morally responsible developers of VR applications.

Funding Funding was provided by Markkula Center for Applied Ethics and Oculus Education Grant.

References

- Ahn, S. J., Bostick, J., Ogle, E., Nowak, K., McGillicuddy, K., & Bailenson, J. N. (2016). Experiencing nature: Embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *Journal of Computer-Mediated Communication*. <https://doi.org/10.1111/jcc4.12173>.
- Anderson, C. A., Keio, A. S., Ochanomizu, N. I., Swing, E. L., Bushman, B. J., Rothstein, H. R., et al. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. *Psychological Bulletin*, *136*(2), 151–173.
- Berger, J. M. (2009). Replicating Milgram: Would people still obey today? *The American Psychologist*, *64*(1), 1–11. <https://doi.org/10.1037/a0010932>.
- Cummings, J., & Bailenson, J. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309.
- Di Nucci, E. (2012). Self-sacrifice and the trolley problem. *Philosophical Psychology*, *26*(5), 662–672.
- Feinberg, J. (1985). *Offense to others*. Oxford: Oxford University Press.
- Felnhofera, A., Kothgassner, O. D., Schmidt, M., Heinzle, A.-K., Beutl, L., Hlavacs, H., et al. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International Journal of Human-Computer Studies*, *82*, 48–56.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Fischer, P., Kastenmüller, A., & Greitemeyer, T. (2009). Media violence and the self: The impact of personalized gaming characters in aggressive video games on aggressive behavior. *Journal of Experimental Social Psychology*. <https://doi.org/10.1016/j.jesp.2009.06.010>.
- Foot, P. (1978). The problem of abortion and the doctrine of double effect. In *Virtues and vices and other essays in moral philosophy*. Oxford: Clarendon Press.
- Fox, J., Bailenson, J., & Ricciardi, T. (2012). Physiological responses to virtual selves and virtual others. *Journal of Cybertherapy and Rehabilitation*, *5*(1), 69–72.
- Frigg, R., & Hartmann, S. Models in Science. *The stanford encyclopedia of philosophy* (Spring 2017 Edition). Edward N. Zalta (ed.), Retrieved from <https://plato.stanford.edu/archives/spr2017/entries/models-science/>.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, *21*(5), 725–740.
- Goldi, P. (2011). Anti-empathy. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 318–330). Oxford: Oxford UP.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*(4), 613–628.
- Handrahan, M. (2016). VR devs call for restraint on horror games and jump scares. Gamesindustry.biz. Retrieved August 9, 2017 from <http://www.gamesindustry.biz/articles/2016-03-14-vr-developers-advise-caution-on-horror-games-and-jump-scares>.
- Herrera, D. C., & Perry, A. (Eds.), (2013). *Ethics and neurodiversity*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Higuera-Trujillo, J. L., López-Tarruella, M., & Llinares, M. C. (2017). Psychological and physiological human responses to simulated and real environments: A comparison between photographs, 360° panoramas, and virtual reality. *Applied Ergonomics*, *65*, 398–409.
- Hudson, L. (2016). It's not real, it's not real, it's not real: Virtual reality and the dangers of 360-degree abuse. *Slate*. Retrieved August 09, 2017 from http://www.slate.com/articles/arts/gaming/2016/05/virtual_reality_can_be_more_amazing_and_ore_terrifying_than_a_traditional.html.
- Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M. J., & Simo, R. N. (2011). A review of the use of psychophysiological methods in game research. *Journal of Gaming & Virtual Worlds*, *3*(3), 181–199.
- Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262–271.
- Krahé, B., & Möller, I. (2010). Longitudinal effects of media violence on aggression and empathy among German adolescents. *Journal of Applied Developmental Psychology*, *31*, 401–409.
- Krohs, U. (2008). How digital computer simulations explain real-world processes. *International Studies in the Philosophy of Science*, *22*(3), 277–292.
- Lejacq, Y. (2013). 'Grand Theft Auto V' torture episode sparks controversy. News, N. B. C. Retrieved January 08, 2017 from <http://www.nbcnews.com/technology/grand-theft-auto-v-torture-episode-sparks-controversy-4B11185864>.
- Liao, M. S., Wiegmann, A., Joshua, A., & Vong, G. (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*, *25*(5), 661–671.
- 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*. <https://doi.org/10.3389/frobt.2016.00003>.
- McLay, R. N., Wood, D. P., Webb-Murphy, J. A., Spira, J. L., Wiederhold, M. D., Pyne, J. M., et al. (2011). A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. *Cyberpsychology, Behavior, and Social Networking*, *14*(4), 223–229.
- Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, *67*, 371–378.
- Morrison, M. (2009). Models, measurement and computer simulation: The changing face of experimentation. *Philosophical Studies*, *143*(1), 33–57.
- Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional 'trolley problem'. *Emotion*, *12*(2), 364–370.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*, 663–685.
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 650.

- Parsons, T. D., Gaggioli, A., & Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*. <https://doi.org/10.3390/brainsci7040042>.
- Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250–261.
- Pastötter, B., Gleixner, S., Neuhauser, T., & Karl-Heinz, T. B. (2013). To push or not to push? Affective influences on moral judgment depend on decision frame. *Cognition*, 126(3), 373–377.
- Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, 9(1), 94–107.
- Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence Teleoperators and Virtual Environments*, 11(1), 68–78.
- Ramirez, E. (2017). Empathy and the limits of thought experiments. *Metaphilosophy*, 48(4), 504–526.
- Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., et al. (2007). Affective interactions using virtual reality: The link between presence and emotions. *Cyberpsychology and Behavior*, 10(1), 46–56.
- Rockstar North, et al. (2013). Grand Theft Auto V. PlayStation 3 & Xbox 360: Rockstar Games.
- Rosenberg, R., Baughman, S., & Bailenson, J. (2013). Virtual superheroes: Using superpowers in virtual reality to encourage prosocial behavior. *PLoS ONE*, 8, e55003. <https://doi.org/10.1371/journal.pone.0055003>.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews: Neuroscience*, 6, 332–339.
- Savage, J., & Yancey, C. (2008). The effects of media violence exposure on criminal aggression: A meta-analysis. *Criminal Justice and Behavior*, 35(6), 772–791.
- Schlenker, B. R., & Forsyth, D. R. (1977). On the ethics of psychological research. *Journal of Experimental Social Psychology*, 13(4), 369–396.
- Shue, H. (1978). Torture. *Philosophy and Public Affairs*, 7, 124–143.
- Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, 8, 426.
- Slater, M., Antley, A., Davison, D., Swapp, D., Guger, C., Barker, C., Pistrang, N., et al. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE*, 1, e39.
- Szabó, T. S., & Liao, S. (2016). The problem of imaginative resistance. In J. Gibson & N. Carroll (Eds.), *The Routledge Companion to Philosophy of Literature* (pp. 405–418). Routledge.
- Valkenburg, P. M. & Peter J. (2013). The differential susceptibility to media effects model. *Journal of Communication*, 63, 221–243.
- Williams, B. (1985). *Ethics and the limits of philosophy*. Cambridge: Harvard University Press.
- Winsberg, E. (2009). A tale of two methods. *Synthese*, 169(3), 575–592.
- Won, A., Bailenson, J., Lee, J., & Lanier, J. (2015). Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20(3), 241–259.