Ethical issues with simulating the Bridge problem in VR

Penultimate draft. Please cite the published version in *Science and Engineering Ethics*.

Abstract: We aim to generate a dilemma for virtual reality-based research that we motivate through an extended case study of Judith Thomson's (1985) Bridge variant of the trolley problem. Though the problem we generate applies more broadly than the Bridge problem, we believe it makes a good exemplar of the kind of case we believe is problematic. First, we argue that simulations of these thought experiments run into a practicality horn that makes it practically impossible to produce them. These problems revolve around concepts that we call "perspectival fidelity" and "context realism." Moral dilemmas that include features present in the Bridge variant will, as a result, be practically impossible to simulate. We also argue that, should we be wrong about the practical impossibility of creating a VR simulation of Bridge, such a simulation must face an ethical horn which renders these simulations ethically impermissible to develop or use. For these reasons, we argue that it is virtually impossible to simulate the bridge problem (and other thought experiments with similar features) both practically and ethically in VR.

Erick Jose Ramirez   *Scott LaBarge*
Santa Clara University   *Santa Clara University*

Philosophers and psychologists interested in moral decision-making have long turned to thought experiments to articulate, gather data on, and defend their theoretical claims.[1] Thought experiments like the Trolley Problem (Foot 1978), and its many variants, have become staple tools in the study of moral psychology as well. The recent development of robust and accessible commercial virtual reality (VR) technologies offer researchers interested in these (and other) questions an exciting new tool for simulating these scenarios. Experiments built around these simulations, many researchers hope, can teach us a great deal about how people would *really* make the difficult moral choices that thought experiments ask them to imagine. They promise, in other words, to shed interesting new light on long-standing philosophical debates over the nature of moral concepts, moral properties, and moral epistemology.

However, designers of such simulations face difficult, and unrecognized, challenges. To see why, consider first many classical philosophical thought experiments. These thought experiments involving moral choice (e.g., the trolley problem, ticking time-bombs, children drowning in fountains, etc.) are typically presented as first-personal. They ask us to imagine what *we ourselves* would do in the situations described by the experiments.[2] Would you pull the lever to divert the trolley? Would you

---

[1] The authors wish to thank Miles Elliott, Mohit Gandhi, Lia Petronio and three anonymous reviewers for many helpful comments while writing this paper.

[2] Although we do not take up the problem here, we believe that experiments that ask us to imagine what it would be like to be *someone else* in such situations encounter their own brand of perspective-taking issues as serious as the ones we discuss here (Ramirez 2017).

torture the terrorist to prevent a nuclear explosion? Would you save the drowning child even if it meant that you would be late to your meeting and ruin your new shoes?[3]

To answer these questions, we must be able to successfully simulate these situations in the imagination and accurately gauge our responses from within those imagined spaces. Thus, one goal for simulations of such thought experiments, especially those in moral psychological research, is to give us insight into precisely the choice that the thought experiment embodies.[4] A successful simulation, on these terms, is one that could give us meaningful insight into how actual people would act or feel or judge when realistically confronted by these choices. Such an experiment, we argue, ought to make an experimental subject feel like she is actually facing the choice scenario posed by the thought experiment with as much realism and sense of immersion built into the experience as possible. Insofar as these experiments aim to draw data (or advance philosophical lines of argument) from the sorts of judgments we make about them, they ought to aim, in other words, for simulating environments with as much ecological validity as possible.

Virtual Reality (VR) systems comprise a diverse set of technologies ranging from rooms in which simulated content is projected onto walls, enclosed simulations used for pilot training, and commercial VR headsets which, when worn, allow a user to fully turn their hands and move around in a virtual space. Because of the immersive nature of VR technologies, researchers have held out hope that VR simulations of these thought experiments can represent a better way of creating such ecologically valid scenarios and improve upon experimental design for studies of moral judgment (Parsons 2015; Kothgassner & Felnhofer 2020; Ramirez 2018; Ramirez & LaBarge 2018).

We claim, however, that in the case of a broad class of first-personal thought experiments, designers face a dilemma with two horns, a "practicality" horn and an "ethical" horn, that together render it virtually impossible to successfully simulate them in VR. The practicality horn aims to show that *structural* design problems inherent in the creation of simulations like Bridge will cause such simulations to lack ecological validity to a degree that thwarts the experimental goal of genuinely capturing what it would be like to face the relevant moral choice. As such, they will fail to count as genuine simulations of the corresponding thought experiments. Should we be wrong about the impossibility of designing such ecologically valid simulations, we claim that the resulting simulations would be nonetheless unethical to expose people to. Such simulations are virtually impossible to create

---

[3] In some instances the question isn't *would you* do x, y, or z but *should*. These more normatively framed thought experiments succumb to similar sorts of problems insofar as they require the subject to engage in an act of perspective-taking in order to respond (Ramirez 2017).

[4] Increasingly, philosophers and psychologists have also appealed to experimental data on such thought experiments in order to support or critique normative and meta-ethical frameworks (Di Nucci 2012; Liao et al. 2012; Pastotter et al. 2013; Huebner and Hauser 2011; Swann et al. 2010; Haidt and Bjorklund 2007). As such, the inference from experimental data to real-world generalization requires a high degree of ecological validity (significant similarity between the experimental situation and the real-world situations it aims to generalize to).

successfully in the sense that they are practically impossible to be built in the first place (the practicality horn) and ethically impermissible to expose people to (the ethical horn) should it be possible to successfully build them. This essay aims to explain the full force of this dilemma and its impact on the creation of these sorts of first personal simulations.

To make our case, we focus on a single classic thought experiment, Thomson's "Fat Man and the Bridge" variant of the classic trolley problem. Hereafter, we refer to this case as "Bridge" (Thomson 1985). To strengthen our claim, we will show how one recent attempt to simulate it (Francis et al. 2016) succumbs to the design problems we highlight. We furthermore claim that the problem with simulations of Bridge are not unique to it but will generalize to attempted simulations of other first-personal thought experiments that bear the same structural features as Bridge.

1. *A Classic Thought Experiment: The Bridge Problem*

Here is the *locus classicus* for Judith Thomson's Bridge thought experiment:

> [Y]ou are standing on a footbridge over the trolley track. You can see a trolley hurtling down the track, out of control. You turn around to see where the trolley is headed, and there are five workmen on the track where it exits from under the footbridge. What to do? Being an expert on trolleys, you know of one certain way to stop an out-of-control trolley: Drop a really heavy weight in its path. But where to find one? It just so happens that standing next to you on the footbridge is a fat man, a really fat man. He is leaning over the railing, watching the trolley; all you have to do is to give him a little shove, and over the railing he will go, onto the track in the path of the trolley. Would it be permissible for you to do this? (Thomson 1985, 1409)

Although Thomson first described Bridge third-personally (1976, 207-208), her canonical (1985) statement is second-personal and the dilemma itself requires a first personal approach.[5] In the intervening years, philosophers and psychologists have followed suit by construing their uses of Bridge as a first-personal moral dilemma (Bruers & Braeckman 2014; Greene et al. 2001). Because of Bridge's significance in philosophical and psychological discussions of moral agency, Bridge and cases like it have long been used as vignettes in moral judgment research. Investigators have, unsurprisingly, begun using VR simulations of the choice scenario Bridge describes to potentially improve upon existing research methods.

As VR technology has become more available, many experimenters have hoped that a turn to VR simulations in their experiments will allow them to gather better, that is, more realistic (ecologically

---

[5] Thanks go to an anonymous reviewer for this clarification.

valid) data than traditional thought experiments allow.[6] "A virtual environment provides the researcher with an ecologically valid platform for presenting dynamic stimuli in a manner that allows for both the veridical control of laboratory measures and the verisimilitude of naturalistic observation of real life situations" (Parsons 2015, 7). In order to see why VR technologies have seemed to hold so much promise, we first must say more about the technology itself and introduce concepts relevant to a discussion of user experiences in virtual worlds.

2. Simulation in Virtual Reality: Some Concepts

Before we discuss specific simulations, we first introduce concepts important to understanding the promises and possible perils of VR technology. We discuss what psychologists call the *presence* that VR users often feel in such simulations and differentiate this feeling from *virtually real experiences* users might have in a simulation. Our central claim in this section is that VR simulations possessing high degrees of what we call *perspectival fidelity* and *context-realism* will tend to produce virtually real experiences. We argue that producing virtually real experiences should be a (if not the) central goal in the design of first-personal simulations that aim to generate ecologically valid data.

One of the distinctive features of VR technology is that it can sometimes make users feel as though they have been transported into a virtual world. Psychologists refer to the subjective experience of being transported into a simulated world as the experience of *presence* (Cummings and Bailenson 2016; Sanchez-Vives and Slater 2005).

Presence helps explain why VR simulations are useful for exploring moral psychology, but we claim that presence alone does not suffice as a design goal for simulations hoping to produce ecologically valid responses in their subjects or users. As many VR games set in fantasy worlds demonstrate, players can feel present in simulated worlds even when the worlds they inhabit contain elements built to discourage their users from experiencing the world as real. By their nature, for example, such simulations are meant to be enjoyed by players. Thus, actions that, in reality, would be negative (e.g. killing a person), must be transformed into something more benign (e.g. an opportunity to boost one's score or to gain a new item). Such simulations may be highly present in the sense that gamers may have the experience that they are really there at the ramparts, fighting the enemy, but it's unlikely that most game designers want players to have genuinely realistic thoughts, feelings, or experiences of these events.

---

[6] Experimenters must always balance various forms of experimental validity when they think through the design of any study. Ecological validity, we argue, is especially important in moral judgment research. If experimenters are interested not just in the moral judgments that are made in the lab environment but wish to generalize from that environment to real-world judgments, then the lab environment should attempt to replicate, to the degree possible, the sorts of situational features that are likely to impact moral judgments in the wild. Ecological validity in particular has thus become the focus of many VR moral psychologists (Slater et al., 2006, Patil et al. 2014, Parsons 2015, Ramirez 2018, Skulmowski et al 2014).

By contrast, we and many other VR researchers have suggested that the experimenter's goal should be to produce simulations in which subjects not only feel present but which also allow them the opportunity to have *virtually real experiences*. Such experiences are simulated experiences that subjects treat *as if* they were real. Virtually real experiences tend to be generated by simulations that not only produce a sense of presence but also contain high degrees of *perspectival fidelity* and *context-realism* (Ramirez & LaBarge 2018).

Perspectival fidelity refers to the degree to which a simulation accurately reproduces the structural perspectival features of human experience. For instance, a simulation which gives the user a point-of-view about six feet above the ground (give or take) is more perspectivally faithful than one that gives her a god's-eye view or one that offers the perspective of a freely moving camera. A simulated perspective with normal color vision, depth of field, and stereoscopy is more perspectivally faithful, a more faithful recreation of human perceptual experience, than a simulated perspective that lacks or substantially alters these features.[7] Not all elements of perspectival fidelity are visual; simulations that include realistic haptic feedback, for instance, are more perspectivally faithful than simulations that don't. Auditory elements that are inexplicable in the context of the simulated world (like disembodied voice-overs, unexplained ambient music, and other non-diegetic meta-sounds) reduce the degree of perspectival fidelity inherent in a simulation.

Despite not being a part of the simulation itself, the hardware a person uses to experience a simulation can impact its degree of perspectival fidelity. For example, the physical sense of wearing a bulky VR headset reduces perspectival fidelity insofar as it intrudes on the experience someone has within a simulation. Other hardware features like the field of view afforded by the headset, the refresh rate of its displays, and so on, have similar impacts on the degree of perspectival fidelity a simulation is experienced as having. Perspectival fidelity is thus multi-dimensional and a matter of degree; different simulations can be more or less perspectivally faithful in different ways. While the *content* of a simulation can be relevant to its perspectival fidelity, a simulation's perspectival fidelity largely depends on a simulation's structural features.[8] The lower the perspectival fidelity a simulation contains, the more likely it is that a person experiencing that simulation will treat their simulated

---

[7] To a degree, perspectival fidelity is user-relative. Individuals vary physically in terms of how they perceive the world, and thus perspectival fidelity conditions will vary between them. Here we speak in statistical terms about typical perspectives. Non-typical subjects will have their own senses of perspectival fidelity that better represent their own typical experiences. Experimenters should of course be sensitive to the range of variability in subjects as they construct their VR simulations to avoid adding confounding variables. Although in many ways user relative, human biology places upper limits on variability for perspectival fidelity. We thank an anonymous reviewer for helping us clarify this point.

[8] Interestingly, it appears increasingly likely that photorealism only weakly affects perspectival fidelity and that behavioral realism (of virtual characters and virtual environments) plays an outsized role (Slater 2006; Zendle, Kudenko, Cairns 2018; Sanchez-Vives & Slater, 2005). Thus, even fairly graphically primitive simulations are capable of being experienced as high in perspectival fidelity if built well. Similarly, VR environments with high resolution photo-realistic environments are not, by themselves, likely to significantly impact how that simulation is perceived by users. We thank an anonymous reviewer for raising the question of photorealistic environments and their effects on virtually real experience.

experiences differently than a real-life counterpart in light of those features. They serve as reminders of the artificiality of the experience and signal to the subject that the experience is not real. Content, however, does also have an important role to play in terms of the degree to which a simulation generates virtually real experiences.

A simulation's content is most likely to impact its degree of a feature we call "context-realism." Context-realism refers to the degree to which the content of a simulation coheres with the rules of the actual world, as understood by the user. "First-person shooter" games, for instance, typically reduce their context-realism by including design elements like "power-ups" that emerge from defeated enemies, or when defeated enemies die unrealistically by disappearing from view or exploding. Simulations with realistic physics are more context-real than simulations without. Simulations that add meta-content to the user's field of view (like health gauges, maps, scores, ammunition gauges, a targeting reticule, etc.) are less context-real than simulations without such content (and for the same reason, such additions also reduce the perspectival fidelity of a simulation). Simulated non-player characters that behave realistically are more context-real than ones that do not.[9] A simulation set in the modern day is more context-real than one set in a distant past or future. Science fiction or fantasy settings are likely to increase the degree to which a subject views the simulation's content as game-like instead of as virtually real.

As with perspectival fidelity, context-realism is a matter of degree. Importantly, since it depends on what the user believes the real world to be like, context-realism is somewhat subjective. For instance, a simulated world that included ghostly characters would, for that reason, be less context-real to someone who doesn't believe in ghosts, while a more ghostly-inclined person might conceivably experience such a simulation as *more* context-real in virtue of the ghosts' presence in the simulated space.

Interestingly, a wide range of VR games set in imaginative fantasy worlds demonstrate that the sense of presence can arise even in a VR simulation with diminished levels of perspectival fidelity or context-realism. Although the feeling of presence may persist in such simulations (in the sense that users may report feeling like they're in the simulated world instead of their homes), we claim that the presence of simulated features that diminish perspectival fidelity and context-realism will typically prevent VR users from having virtually real experiences in such simulations. They will be unlikely to treat the experiences they have in such simulated worlds as if they were real experiences. In our terms, a virtually real experience of violently killing other human beings would (hopefully) turn out to not be

---

[9] As previously noted, non-player characters that behave realistically appears to be a much more important element of context-realism than designing a photo-realistic environment (Zendle, Kudenko, Cairns 2018)

very welcome to most players.[10] Some experiences we positively do not want to experience as real, and for that very reason designers might consciously choose to make their simulations less perspectivally faithful or context-real than they might otherwise be.

In the case of the thought experiments researchers aim to simulate, however, there are good reasons to think that the production of simulations that generate virtually real experiences (as opposed to a mere sense of presence) should be the goal. If researchers want these VR experiments to accurately collect information about what people would really do, feel, and think in these imagined thought experiment scenarios, then they need their subjects to treat their experimental experiences as if they were real to the highest degree possible.

For instance, researchers would not want subjects in a VR Bridge simulation to treat their experiences as artificial or gamelike. Data generated in such an experiment simply wouldn't be useful to answering questions about moral judgments in the relevant way (though they may be useful in other ways) and wouldn't shed light on real-time real-world moral judgments.[11] Researchers should want their subjects to feel like they were making genuine choices in the context of a world as much like the real world as possible, and with consequences as much like real-world consequences as can reasonably be managed (i.e. they should want to create ecologically valid simulations). Insofar as producing virtually real experiences is a matter of designing simulations high in perspectival fidelity and context-realism, we should seek to design our simulations with those qualities in mind. Existing VR simulations, however, have largely ignored these concerns.

3. Experimental problems with perspectival fidelity and context-realism

If the point of using VR simulations of Bridge is to get information about what people would actually do, then simulations of Bridge need to generate virtually real experiences to give experimenters access to that kind of information.[12] In this section we argue that researchers aiming to simulate Bridge in VR

---

[10] Indeed there is some evidence that a simulation's structural design can impact the contours of a subject's moral judgments about the virtual actions they engage in and that these differences may generate what has been referred to as "The Gamer's Dilemma" (Ali 2015, Luck 2009, Ramirez 2020).

[11] Similarly, although written prompts and gamelike simulations may trigger dual process judgment pathways (Greene et al. 2001), it remains to be seen whether such pathways are triggered in the same way in ecologically valid scenarios as they are in written and gamelike scenarios. This is largely an empirical issue about which data is scarce. We thank an anonymous reviewer for raising the question of emotion in moral judgment studies.

[12] Simulations with different aims, however, could avoid our dilemma entirely. For example, Ahn et al. (2016) investigated whether embodying subjects with animal bodies affected their perceptions of environmental issues. In such a case, ecological validity is, arguably, impossible (there does not exist a context where we could become lobsters or cows). These studies thus have low degrees of perspectival fidelity and context-realism and probably didn't generate virtually real experiences in subjects (even though they clearly generated meaningful experiences). The ersatz experiences of animal embodiment may, however, have relevant forward-looking behavioral effects on subjects and such an experiment may neither want, nor need, to create virtually real experiences in its subjects in order to affect behavior in an experimentally relevant way.

face a dilemma involving what we believe are insurmountable practical design problems with creating such simulations. Though we believe that these problems are genuinely intractable (and hence explain one way in which it is impossible to simulate these thought experiments), we accept that our imagination is limited and we could very well be wrong about this impossibility. In the next section, we take seriously this possibility and argue that even if this were to turn out to be the case, the resulting simulations would have ethical problems that ought to render them morally impermissible to expose people to (and hence represent yet another way in which such simulations would be virtually impossible to create).

At its heart, Bridge asks for a moral judgment: would we push (or not push) a man onto the tracks of a runaway trolley, or, alternatively, how would we judge another person's actions when they acted in either way? Any simulation of choices like these must accurately recreate the choice scenario experimentally or run into problems with ecological validity (Ramirez 2019). That is, a successful simulation of Bridge needs to provide its subject with virtually real experiences of being on a bridge with a runaway trolley threatening others below, it needs to provide the experience of having the idea that pushing the man onto the tracks could realistically save the five down below, and it needs to include a sense of genuine time-sensitivity to the decision (so much so that asking the man to sacrifice himself is not an option given the circumstances).

The first problems when simulating Bridge involve design issues concerning perspectival fidelity and context-realism. A simulation of Bridge, if it aims to get feedback about what people would *actually do, think, judge, or feel* in a real-life Bridge scenario needs to accurately recreate the conscious (doxastic) and unconscious (subdoxastic) elements of the original scenario (Ramirez 2017). Recent attempts to simulate Bridge help to show why this is especially difficult to do given the difficulty of creating a perspectively faithful and context-real simulation of first personal moral dilemmas. For example, one recent attempt to recreate Bridge using VR utilized the following protocol:

> The landscape…was kept neutral with hills in the background and a neutral "skybox"…. Verbal instructions played during the 3D scenario and specific instructions were given prior to the experimental task, explaining that this task involved a joystick but that participants would be given a choice about whether they wanted to interact with the virtual object or not…. After 30 seconds, verbal instructions informed participants that a trolley car was approaching ("*Look behind you, a train is coming.*") After a further 25 seconds, a second verbal dialogue then followed ("*Hey I am too far away but if you want to save the people you could push the large person on to the tracks and derail the train. If you're going to push him, do it now, but it is your choice.*") (Francis et al. 2016, 5/22)[13]

---

[13] Original emphasis

We want to highlight several issues with this simulation of Bridge in order to make a general point about the possibility of simulating first-personal thought experiments more generally. In doing this, we will pause to note the design choices these experimenters made when creating their simulation and how those choices worked to impact the simulation's degree of context-realism and perspectival fidelity.

Before beginning, we pause to appreciate that methodological design choices always represent compromises between experimental precision (in the ideal, only the independent variable of an experiment should differ between instances) and ecological validity (natural environments contain many variables which might catch a subject's attention in a way that isn't experimentally ideal). Though some of the problems we describe below can be resolved or diminished (in other words, we believe it possible to create more perspectively faithful and context-real simulations than the one under discussion), other design problems with this simulation would be an inherent part of any simulation of Bridge and ultimately make such simulations unfaithful reproductions of the original thought experiment. These problems, for clear reasons, are more serious.

To begin, we argue that the experimenters' choice to leave the simulated world barren diminishes its context-realism. The footbridge exists in a completely featureless world. Indeed, the footbridge itself is not naturally connected to the barren world it is set within. No footpaths exist on either end of the bridge to indicate that it is a natural part of its environment or that it serves a functional purpose. These design features thus work to deliver a decidedly context-unreal environment to its subjects. Subjects are also not given an explanation as to how they ended up on the bridge, why there is a man on the bridge with them, or their relationship to the man. Additionally, no explanation is offered why the other individuals have become stuck on the tracks. They are simply, inexplicably, there. Because such explanations are natural parts of real-world moral dilemmas (i.e., real-world moral judgments are situated in narratively rich contexts) (Zagal 2009), leaving out these elements diminishes the simulation's degree of context-realism and thus decreases the likelihood of the simulation generating ecologically valid decisions.

Additionally, the man on the bridge does not respond in any way to the subject's existence, nor does he respond at all to being shoved off the bridge. This ought to be puzzling to subjects and would, we claim, threaten to shift how a subject in such an experiment experiences the moral dilemma itself. Failure to model naturalistic human behaviors in this way can dramatically impact how a subject understands the moral status of virtual agents (Slater et.al. 2006).[14] This absence of realistic agency, we

---

[14] The lack of reasons-responsiveness on the part of the man on the bridge works to diminish the context-realism of the scenario and thus diminish the experiment's ecological validity. In the aforementioned study by Slater et al., Slater replicated, in virtual reality, the results of Stanley Milgram's (1963) obedience studies - showing that it is possible for subjects to treat virtual experiences as if they were real. Even though Slater's environment was graphically primitive, it was still able to generate virtually real experiences, providing further evidence that simply creating more photorealistic environments is unlikely to affect ecological validity without also increasing context-realism and perspectival fidelity.

argue, radically diminishes the simulation's context-realism (Zendle, Kudenko, & Cairns 2018).[15] When subjects in this experiment chose to push the man onto the tracks, these features of the simulation also make it more likely that subjects treat their experience of their decision in virtually unreal (game-like) ways.

Other features of the simulation work to diminish *both* its context-realism and perspectival fidelity. Consider, for example, the decision to use a joystick for locomotion. Subjects are apt to associate joysticks with gaming (a real potential confound), and this choice thus diminishes the context-realism of the scenario. Beyond that, using a joystick (instead of naturalistic bodily movements) to push the man onto the tracks would also work to diminish the perspectival fidelity of the simulation by poorly recreating the embodied experiences of actually moving and of pushing called for by the original thought experiment (Bianchi-Berthouze, Kim, & Patel 2007). These features, we believe, could arguably be addressed in future VR simulations of Bridge in the sense that experimenters can choose to design their experiment such that subjects have a rationale for being on a context-real bridge and are present for a realistic sort of disaster that might leave five unlucky people stuck on the tracks. They may also choose to design their simulation such that subjects must physically move their bodies and use their arms to push the virtual man off the bridge and onto the tracks.

However, the most problematic design feature of the simulation, the one that we believe would cause trouble for *any* simulation of Bridge, is how Francis et al. (2016) delivered crucial information to their subjects. Key to Bridge is the fact that the subject is supposed to *know* that pushing the man onto the tracks will stop the trolley. This point is made emphatically in philosophical (Thomson 1985) and experimental (Greene et al. 2001) versions of Bridge.

> Being an expert on trolleys, you know of one certain way to stop an out-of-control trolley: Drop a really heavy weight in its path. But where to find one? It just so happens that standing next to you on the footbridge is a fat man, a really fat man. (Thomson, 1985, 1409)

This information, and the fact that it is meant to come *from the subject*, is crucial to the decision-making structure of Bridge. It is difficult — indeed, we claim it's practically impossible — to deliver this information to subjects without either changing the dilemma itself or significantly diminishing the simulation's perspectival fidelity and context-realism and thus diminish the ecological validity of the simulation. Because any simulation of Bridge would have to confront this problem, we believe it is especially instructive to show not only how it goes awry in the present study but also why this problem will generalize to other simulations. We argue that the design problem posed by this feature of Bridge (having subjects realize, on their own, that they could push the man to save the five),

---

[15] It may also diminish, for related reasons, the simulation's degree of perspectival fidelity.

cannot be simulated in a perspectively faithful or context-real way while *simultaneously* being a faithful representation of Thomson's original scenario.[16]

In Francis' simulation, a non-diegetic voice-over alerts subjects to the presence of the trolley and also tells them that they could, with certainty, push the man onto the path of the trolley to save those on the tracks. The design decision to use non-diegetic voice-over radically diminishes both the perspectival fidelity and context-realism of the simulation. The non-diegetic character of the voice-over leaves it entirely up to individual subjects to interpret its source and its relevance. Does it, for example, represent the voice of God? Are subjects receiving audio instructions from a nearby engineer? Are subjects supposed to understand the voice as a particularly strange sort of moralizing hallucination? Could it be the voice of one of the researchers speaking to them from beyond the simulation?

Arguably, any of these interpretations of the source of the non-diegetic audio cue will work to diminish the subject's responsibility. For example, if the voice-over is interpreted as being that of a nearby engineer, the decision to push the man onto the tracks can be partially attributed to the engineer for suggesting it in the first place; if attributed to God, the voice arguably becomes the determiner of moral content (if God is suggesting I push the man then it can't be wrong!). If the voice-over is understood as a hallucination, this may cause the subject to question the nature of the simulation itself. Worse still, if subjects interpret the audio as the voice of one of the experimenters, then we face issues with subject behavior in the face of demand characteristics. If what we want is to create a simulation of Thomson's (1985) thought experiment, it's crucial that subjects generate, on their own, the belief that the man on the bridge *can* be used to stop the trolley; delivering this information externally (diegetically or non-diegetically) won't do without fundamentally altering the nature of the experiment. However, it is difficult to see how to construct a perspectively faithful, context-real, simulation of Bridge that manages to do this.

Non-diegetic voice-over and other such design choices effectively break simulations of this kind. Because such design choices diminish perspectival fidelity and context-realism, subjects in such simulations will be less likely to have virtually-real experiences: they are not thinking, feeling, or judging as they would in a realistic, ecologically valid simulation of Bridge. Instead, a subject's decision is more likely to represent either an exploration of the options allowed by the simulation (as in a game), their falling prey to demand characteristics in the experiment, or an attempt by subjects to predict their own decisions about what they would or ought to do in the simulative circumstances. None of these

---

[16] One can create an infinite variety of moral dilemmas involving bridges, trolleys, and sacrifices but to recreate Judith Thomson's Bridge experiment one must keep the essential details of the simulation steady. As we say below, it seems to us (at the very least) that the *source* of information about the large man is a relevant moral variable (i.e., that it matters where this knowledge comes from) and that changing this variable alters the experiment.

options accurately represents the descriptive or normative judgments they would make in ecologically valid versions of Bridge.[17]

Importantly, *any* attempt to simulate Bridge must confront these design problems. Although it is possible to more genuinely integrate the bridge into a less barren world and to script more realistic responses for the man on the bridge and for those trapped down below, it is practically impossible, we argue, to design a simulation of Bridge that manages to get subjects to spontaneously decide that they can push the man onto the tracks in order to derail the train and save the five people below. Attempts to deliver this information will usually (we think *always*) introduce elements into a simulation that diminish its perspectival fidelity and context-realism and keep it from being an ecologically valid simulation of Thomson's original thought experiment.

Simulations of Bridge (and other thought experiments that include similar features) are not possible to successfully create because it seems impossible to design a perspectivally faithful and context-real simulation of Bridge in which subjects generate, on their own, the belief that the man can be pushed in order to save those below. Any simulation that aims to successfully model Bridge will have to construct its scenario so as to convey the relevant information; however, attempts to introduce this information in an ecologically valid way are likely to fail. We thus believe that it is virtually impossible to create a successful, ecologically valid, simulation of Bridge.

Even if, however, we turn out to be wrong about the practical possibility of creating such a simulation, we believe another, distinctly moral, problem would confront the creators of such simulations. Call the concerns we have just elaborated about the practical possibility of creating a successful Bridge simulation the *practicality horn* of a dilemma. We think it unlikely anyone could overcome that horn. However, we're also aware that our imaginations are limited and that some of our empirical conjectures may turn out to be false. History is replete with examples of the supposedly impossible being possible after all. Should this turn out to be the case, we argue that researchers still run afoul of an *ethical horn* of this dilemma. As we will argue, the ethical problems that arise if researchers are able to surmount the practicality horn would make it ethically impermissible to expose subjects to such simulations. Simulations like Bridge are thus practically impossible to design, or wrong to use ethically if we could in fact design them.

4. Moral Problems

---

[17] Demand characteristics, for example, occur when subjects of an experiment aim to please (or frustrate) a researcher under the belief that they know what the experiment itself is investigating. Although subjects can do this consciously, their behavior can also be affected by demand characteristics unconsciously. For this reason, it's important that Francis et al.'s decision to introduce the crucial information regarding the man not, by itself, confound the scenario.

Though we believe that the practicality horn of the Bridge dilemma is probably not surmountable, clearing this hurdle would present researchers with additional and substantial ethical problems. In particular, we argue that two ethical issues arise for any ecologically valid VR simulation of Bridge (and simulations like it). First, we raise a concern about the trauma such a simulation could expose its subjects to. Second, putting trauma to one side, we argue that there are special moral harms done to subjects by simulations whose design inculcates immoral traits in its users.

There is good evidence that VR experiences can be harmful to their subjects. Specifically, virtually real VR experiences of traumatizing events are likely to be as (or almost as) harmful as their real-life equivalents. For those who doubt that VR experiences, even virtually real ones, can be harmful, consider the case of Mel Slater et al.'s (2006) replication of Stanley Milgram's (1963) obedience studies. Slater's group created a VR simulation that we believe was high in perspectival fidelity and context-realism and thus generated ecologically valid VR experimental conditions (Ramirez 2019). As a result, not only did Slater's group replicate Milgram's original results, the study also seems to have unfortunately generated an ethically troubling degree of trauma in its subjects. This was especially true when, much as in Milgram's original experiments, subjects had to deliver progressively intense shocks to a virtual learner:

> …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)

Virtually real simulations of thought experiments like Bridge are similarly poised to generate real trauma. In Bridge specifically, virtually real experiences of pushing the man on the bridge to his death, or of watching the five on the tracks be crushed by the trolley, could be traumatizing to experience, for obvious reasons. Indeed, we should expect this sort of response given our claims about the interactions between perspectival fidelity, context-realism, and virtually real experience.[18]

Other thought experiments are also likely to traumatize subjects in this way. Imagine, for example, a virtually real experience of torturing someone in a ticking time-bomb scenario (Shue 1978), or of having virtually real experiences of Joel Feinberg's (1985) "Ride on the Bus" cases.[19] In our view, if it

---

[18] As another datapoint, consider the incredible success of virtual reality exposure therapy (VRET). Such therapies make use of context-real and perspectivally faithful simulations to treat both post-traumatic stress disorder and specific phobias. That such simulations can have such success, we claim, is best explained by the fact that such simulations are ecologically valid and thus generate virtually real experiences which, as with traditional exposure therapies, can generate realistic responses in patients. For more, see (Ramirez 2020, McLay et.al. 2011; Parsons and Rizzo, 2007)

[19] Such cases include witnessing bestiality, consensual sex, mutual vomit-eating, coprophagia, and a group smashing a corpse's face with a hammer, all while you are stuck on a public bus.

would be wrong to subject a person to a particular real-world experience, then it would be wrong to subject someone to a virtually-real analogue of that experience.[20] It would therefore be wrong to subject a person to virtually real experiences like those in these thought experiments without their consent due to the likelihood that such experiences could harm the person having them. However, because knowledge of an experiment's content and risk is likely to influence a subject's behavior in the experiment, consent in potentially traumatic experiments (especially those that deliver little to no benefit to the experimental subject) will be difficult to secure and such experiments are unlikely to receive approval from Institutional Review Boards.

At this point, a critic might pause and wonder whether we are not raising more (mere) practical concerns. After all, similar experiences can be traumatic to one subject and not to another, and thus it is possible that some might not be traumatized by a virtually real simulation of Bridge (assuming it's possible to construct such simulations in the first place). If such people exist, then experimenters might be able to find a way to screen against subjects who might be traumatized by their experiment and only expose hardier subjects to simulations of Bridge. While, on the one hand, this appears to defuse at least one kind of ethical issue with such experiments, this response by nature reintroduces problems of experimental generalizability; without a random population sample to conduct research on, such experiments inherently limit what experimenters could learn about moral psychology only to such hardier folk. On the other hand, even if issues of virtual trauma could be addressed, another ethical concern would remain for simulations like Bridge.

Even if the practicality horn of our dilemma can be surmounted, and even if we could be confident that subjects would not experience virtual trauma (two things we think very unlikely), we argue that it would still be wrong to create and use these simulations. To see why, we need to go back to what we believe is an important feature of Bridge (and countless other thought experiments). One essential element of the moral dilemma created in Bridge is that the subject must see the man on the bridge as a potential solution to a trolley problem. We are, for reasons not explained in either the original thought experiment or VR simulations of it, not able to communicate with this man, and hence cannot get his consent to push him. We must, instead, simply see him as a possible way to save five lives. This way of seeing people, we argue, is immoral on most (though of course not all) moral frameworks. If we are right about this, then a simulation that encourages or requires users to see others in immoral ways (that is, make them more likely to see people around them as instrumental solutions to problems) is thereby arguably an immoral simulation.

Consider just how unnatural the Bridge scenario is. We posit that most subjects would never see pushing the man onto the path of the trolley as a solution to the problem without external prompting.

---

[20] The wrongness would increase as a simulation's degrees of perspectival fidelity and context-realism increase.

[21] We are not used to seeing other people (and their bodies) as instrumental solutions to problems, and, when prompted to do so, subjects usually recoil (Greene et al. 2001).[22] Although this common reaction doesn't by itself show that it's immoral to push the man without his consent, we believe that it lends some evidence to the position that *instrumentally viewing the man* as a solution is generally considered wrong. We argue that becoming the type of person who views others as instrumental solutions to moral problems is harmful not only to those who might be used by such a person but *to the person themselves*. We argue that any simulation that makes us more likely to see others in such a way is therefore harming us in ways that extend beyond the subjective trauma the experience itself might cause.[23]

In what follows, we argue that three major normative frameworks (virtue theoretic, deontological, and consequentialist) lend prima facie support to our position. What we offer is merely a sketch of what such arguments would look like and thus is necessarily brief in presentation and cannot definitively be said to prove that these simulations are unethical. However, we believe that a discussion like this has been missing from present discussions about the harms of VR simulations and we hope that these arguments help initiate that very discussion. By their nature, moral frameworks are sensitive to particulars, and thus our discussion of how a specific moral framework would treat the inculcation of personality traits that make someone more likely to view their fellows instrumentally should be subject to what we hope is intense debate and disagreement, including with the arguments we offer here. Our own arguments are not the final statement on these matters.

Virtue theoretic frameworks, for example, often aim to tell us what sorts of persons it would be good for us to *become*. Such persons will work to develop virtuous, i.e., good, character traits (complex combinations of behavioral, psychological, and emotional dispositions to act), while avoiding vices. From a virtue theoretic standpoint, simulations like the ones under discussion seem *prima facie*

---

[21] This, however, is a VR experiment we *could* do: create a perspectivally faithful and context-real simulation of a runaway trolley with a man on a bridge overlooking the tracks that, if pushed, would stop the tracks. Such a simulation would need to avoid prompting (of the sort we've discussed) to suggest that pushing the man off the bridge could save the five people below. We could then see how many subjects generate that idea on their own. Our prediction: few subjects would generate that idea and even fewer would push the man onto the tracks to see if he would in fact stop the train.

[22] Though Joshua Greene's (2001) research provides empirical support to the view that these "personal" moral dilemmas tend to be treated this way, our own experience asking students about Bridge coheres with his more rigorous data. Bridge, we've found, can nearly always be counted on to generate lively discussion of how the thought of pushing the man would ever occur to someone in the first place! Others who teach ethical theory can draw upon their own experiences here to see if they overlap with ours.

[23] In making this argument, we don't intend close off discussion on the ethics of instrumentally viewing others. Our point is that such a discussion is currently lacking in the philosophical and psychological literature regarding the use of VR simulations with human subjects. As such, our hope is that these very questions and arguments weigh on the minds of members of Institutional Review Boards who must approve the use of human subjects for research. We think considerations not only of virtual trauma but also of character change ought to form a part of any assessment of the relative risks, harms, and benefits to human subjects in VR simulations. Virtually real simulations thus deserve a special level of ethical scrutiny that less virtually real simulations would not. Not every simulation of Bridge would thus be intrinsically problematic though we argue here that virtually real formulations encounter the dilemma we have presented.

harmful in the sense that they appear to inculcate vice in subjects. Such simulations make their subjects less likely to act virtuously in their real-world moral affairs by developing psychological and behavioral dispositions that encourage them to see others instrumentally. Such a perspective not only fails to accord others due respect, it also appears to make them less likely to develop beneficent, empathic, or conscientious traits. To the degree that character is formed in part by behavior, and to the degree that such simulations produce virtually real experiences in subjects, such simulations would therefore work to create what are arguably vicious character traits in their subjects.[24]

Similarly, the inculcation of an instrumentalist view of persons would appear to run afoul of Kant's second formulation of the Categorical Imperative to "[a]ct in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means" (Kant 1785/1981, 36). Both Thomson's (1985) original and Francis et al.'s (2016) VR versions of Bridge treat the man on the bridge purely instrumentally, without regard to his own well-being or agency. In fact, these features of the simulation seem to be essential to it (that is, it's important that the only way to save the five on the tracks involves pushing a rational agent into its path without treating them as persons with whom we can communicate or ask for rational consent).[25][26] Such a simulation, were it to successfully lead someone to view others as mere means to the end of saving five lives, would, in doing so, encourage its subjects to fail to manifest the goodwill we have a duty to manifest to our fellow legislators in the Kingdom of Ends.

We believe that even many consequentialists could side against the deployment of ecologically valid simulations of Bridge. Why? From a consequentialist perspective, we believe that the long-term consequences that might arise from developing and using such a simulation would factor heavily in any analysis of its moral permissibility. Consequentialism, like any moral framework, admits of a large number of subspecies and, especially with act consequentialist formulations of the principle of utility, can trend toward forms of moral particularism that make generalizations about ethics difficult. Having said that, we yet believe that many variations of consequentialism, its rule-bound versions most

---

[24] Because VR is used not only for virtual reality exposure therapy but also to train surgeons (Aïm et al. 2016), soldiers (Lele 2013), and for pedagogical training (Jensen & Konradsen 2018), there is at least some evidence that one's behavior in a context-real perspectively faithful VR simulation can affect real-world character traits.

[25] One feature our short analysis brings out is that different moral frameworks might find different classes of virtually real thought experiments unethical in the sort of way we are arguing for here. For example, the deontologist may argue that the same features that make Bridge morally impermissible would make the traditional Trolley Problem equally problematic. One feature of Bridge that makes it an especially suitable target to focus on is that we believe many moral frameworks will converge on the view that its effects on users render it morally impermissible to use and hence highlight the issue we most want to bring to bear here.

[26] Because any virtually real simulation of Bridge requires that we view the consent of the man as irrelevant, or at least as not necessary, to the moral legitimacy of the outcome and because a perspectively faithful and context-real simulation of Bridge would need to have us *form* that thought on our own, it's likely that contractarian approaches to ethics (which place a heavy emphasis on the value of consent) would also disapprove of the wide deployment of such simulations.

especially, suggest that the creation of agents willing to view others as mere means is likely to lead to worse overall consequences, especially for those utilitarians who defend prioritizing special ties.[27]

Although by no means certain, it seems plausible that a world where agents are trained to view one another in instrumental ways would result in the generation of less overall long-term welfare than one where individuals don't view one another in that way. If special ties have intrinsic value, as some utilitarians contend, then the inculcation of traits that encourage instrumental treatment of others would erode our access to such values by making these relationships more difficult.

As such, a consequentialist analysis ought to focus not only the specific "kill one to save five" immediate consequences of Bridge (and the Trolley Problem generally) but also on the long-term consequences that would result from using a simulation whose effect would be to increase the number of people willing to see their fellow citizens as people whose consent for self-sacrifice was not relevant. It seems plausible to us (for what it's worth) that such long-term consequences would speak against using simulations like Bridge, though the final analysis would of course depend on one's specific accounting and weighting of the relative worth of moral variables. By its nature, consequentialism will turn on specific formulations not only of the goods to be produced but also on the actual (or expected) outcomes of a given action (or rule). Our intention here is to suggest that an argument against the implementation of VR simulations of Bridge can be grounded within this moral framework as well.

Although any ethical analysis of Bridge's effects on users will depend in part on its actual outcomes (and their severity), we believe that our first pass arguments here are not implausible. We thus think that it is *prima facie* likely that the ethical horn of our dilemma has real teeth (even if we think the practicality horn is likely insurmountable). Thus, even if it were possible to solve the practicality horn of our dilemma and thereby create ecologically valid VR simulations of Bridge, ethical considerations would nonetheless make it impermissible to deploy them. On the one hand, virtually real experiences of Bridge are very likely to cause unacceptable trauma to subjects. On the other hand, even if such trauma could be avoided, simulations like Bridge require its users to see others as instrumental in a way that, arguably, many of the most frequently invoked and defended moral frameworks would speak against.

5. Conclusion

---

[27] Richard Arneson (2003), for example, has argued that seemingly non-consequentialist values, like special ties, can, and should, be included within the context of act-consequentialism: "[a]ll else being the same, it might be intrinsically better that parents care for their own children rather than that they bestow the identical care on other children. Relationships of friendship might be intrinsically good, so that all else being the same, bringing it about that people become friends makes their lives intrinsically better, and friends helping friends might be intrinsically more productive of value than strangers helping strangers" (383). We believe a similar argument can be rallied to show that such consequentialists would find the VR simulation of these thought experiments objectionable.

Though much remains to be said, we hope to have motivated the claim that an ecologically valid, perspectivally faithful, context-real, and simultaneously ethically acceptable simulation of Bridge would be virtually impossible to produce.

Even if it were possible to overcome what we call the practicality horn of the dilemma, which we think insurmountable for simulations relevantly like Bridge, successful simulations of Bridge run into moral problems that render it morally unacceptable to deploy.

Moreover, while we focus on Bridge, the problems we highlight arise for other thought experiments as well. Any philosopher or psychologist who wishes to employ VR simulations of scenarios like Bridge to study moral decision-making will need to confront these issues as they attempt to design VR simulations of classic moral thought experiments.

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*, 21 (6), 399-419

Aïm, F., Lonjon G., Hannouche, D., & Nizard, R. 2016. Effectiveness of virtual reality training in orthopaedic surgery. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 32 (1), 224-232

Ali, R. 2015. A new solution to the gamer's dilemma. *Ethics and Information Technology*, 17: 267-274

Arneson, R. 2003. Consequentialism vs. special-ties partiality. *The Monist*, 86 (3), 382-401

Bianchi-Berthouze N., Kim W.W., Patel D. 2007. Does body movement engage you more in digital game play? and why?. In: Paiva A.C.R., Prada R., Picard R.W. (eds) *Affective Computing and Intelligent Interaction. ACII 2007. Lecture Notes in Computer Science*, vol 4738. Springer, Berlin, Heidelberg

Bruers, S. & Braeckman, J. 2014. A review and systematization of the trolley problem. *Philosophia*, 42, 251-269

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), 662–72.

Feinberg, J. 1985. *Offense to Others*. Oxford: Oxford University Press

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

Francis, K.B., Howard, C., Howard, I.S., Gummerum, M., Ganis, G., Anderson, G., Terbeck, S. 2016. Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE*, DOI:10.1371/journal.pone.0164374

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. 2001. An fMRI investigation of emotional engagement in moral Judgment. *Science*, 293 (5537), 2105-2108 DOI: 10.1126/science.1062872

Haidt, J., and F. Bjorklund. 2007. Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong, (ed.), *Moral Psychology, vol. 2: The Cognitive Science of Morality*, 181–217. Cambridge, Mass.: MIT Press.

Huebner, B., and M. Hauser. 2011. Moral judgments about altruistic self-sacrifice: When philosophical and folk intuitions clash. *Philosophical Psychology*, 24 (1), 73–94.

Jensen, L. & Konradsen, F. 2018. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23 (4), 1515-1529

Kant, I. 1785/1981. *Grounding for the metaphysics of morals*. James. W. Ellington (translator). Hackett Publishing Company: Indianapolis.

Kothgassner, O.D. & Felnhofer, A. 2020. Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? Annals of the International Communication Association. DOI: 10.1080/23808985.2020.1792790

Lele, A. 2013. Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4 (1), 17-26

Liao, M. S., A. Wiegmann, A. Joshua, and G. Vong. 2012. Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*, 25 (5), 661–71.

Luck, M. 2009. The Gamer's Dilemma: An analysis of the arguments for the moral distinction between virtual murder and virtual paedophilia. *Ethics and Information Technology*, 11: 31–36

Milgram, S. 1963. Behavioral study of obedience. *Journal of Abnormal and Social Psychology 67*, 371–378

McLay, R.N., Wood, D.P., Webb-Murphy, J.A., Spira, J.L., Wiederhold, M.D., Pyne, J.M., Wiederhold, B.K. 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), 223-9

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., and 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

Pastotter, B., S. Gleixner, T. Neuhauser, and T. B. Karl-Heinz. 2013. To push or not to push? Affective influences on moral judgment depend on decision frame. *Cognition*, 126 (3), 373–77.

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

Ramirez, E. 2017. Empathy and the limits of thought experiments. *Metaphilosophy*, 48 (4), 504-526

Ramirez, E. 2019. Ecological and ethical issues in virtual reality research: A call for increased scrutiny. *Philosophical Psychology*, 32 (2), 211-233

Ramirez, E. & LaBarge S. 2018. Real moral problems in the use of virtual reality. *Ethics and Information Technology*, 20 (4), 249-263

Ramirez, E. 2020. How to (dis)solve the Gamer's Dilemma. *Ethical Theory and Moral Practice*, 23, 141-161

Sanchez-Vives, M.V., & Slater, M. 2005. From presence to consciousness through virtual reality. *Nature Reviews: Neuroscience*, 6, 332-339

Shue, H. 1978. Torture. *Philosophy and Public Affairs*, 7,124–43

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*. DOI: 10.3389/fnbeh.2014.00426

Slater, M., Antley A., Davison D., Swapp D., Guger C., Barker C., Pistrang N., & Sanchez-Vives M.V. 2006. A Virtual Reprise of the Stanley Milgram Obedience Experiments. *PLoS ONE*, 1, e39

Thomson, J. 1976. Killing, letting die, and the trolley problem. *The Monist*, 59, 204-217

Thomson, J. 1985. The trolley problem. *Yale Law Journal*, 94 (6), 1395-1415

Zagal, J. P. 2009. Ethically notable videogames: Moral dilemmas and gameplay. *Proceedings of the Digital Interactive Games Research Association Conference (DiGRA)*. Digital Interactive Games Research Association. Available at http://www.digra.org/wpcontent/uploads/digital-library/09287.13336.pdf. Accessed July 15, 2018

Zendle, D., Kudenko, D., Cairns, P. 2018. Behavioural realism and the activation of aggressive concepts in violent video games. *Entertainment Computing*, 24, 21-29

Zimbardo, P. 2004. A situationist perspective on the psychology of evil: Understanding how good people are turned into predators. In *The Social Psychology of Good and Evil*, edited by A. G. Miller, 21–50. New York: Guilford Press