Doing Good with Virtual Reality:

The Ethics of Using Virtual Simulations for Improving Human Morality

Abstract: Much of the excitement and concern with virtual reality (VR) has to do with the impact of virtual experiences on our moral conduct in the "real world". VR technologies offer vivid simulations that may impact prosocial dispositions and abilities or emotions related to morality. Whereas some experiences could facilitate particular moral behaviors, VR could also inculcate bad moral habits or lead to the surreptitious development of nefarious moral traits. In this chapter, I offer an overview of the ethical debate about using VR to improve morality. I start by clarifying the rationale of using VR for good, drawing on moral enhancement literature, branches of the science of morality, and the specific potentials of this technology. Then, I will briefly focus on three prominent domains of socio-moral improvement: the use of VR for enhancing empathy, reducing implicit biases, and improving pro-environmental behavior. After that, I shall consider three ethical objections to the enthusiasm for using VR in order to improve human morality. Finally, I will recapitulate the main points of the chapter and provide a few concluding thoughts on future avenues of discussion on how VR can make us act (morally) better.

Keywords: Virtual Reality Ethics; Ethics of Technology; Virtual Morality; Moral Enhancement; Moral psychology.

1. INTRODUCTION

Can Virtual Reality (VR) become a force for good[1]? The possibility that virtual experiences can lead to positive social change in the non-virtual world is exciting. VR

---

[1] The rhetoric of 'good' has been at the forefront of various VR projects. For instance, the company Oculus, which belongs to Facebook, now rebranded as Meta, runs the program "VR for Good". This specific program promotes and funds immersive journalism aiming to create progressive social impact. (See https://about.meta.com/community/vr-for-good/; last access 12th May 2023). Various immersive films for

comprises a set of technologies that enable vivid, interactive, and immersive experiences in digitally created environments (Slater & Sanchez-Vives, 2016; Bailenson, 2018).[2] If experiences in virtual environments could induce beneficial attitudinal and behavioral changes in the "real physical world", these technologies could become important allies in achieving significant socio-moral goals. Still, assuming this promising possibility could mean accepting its less pleasant counterpart—that VR technologies can promote bad moral traits and lead to ethically problematic dispositions in users.

In this chapter, I philosophically address the impact of VR on human morality. This issue is crucial from an ethical perspective in part because the (un)desirability of many virtual practices must be judged in terms of their benevolent or malevolent influence on the development of users' moral character and conduct (Cotton, 2021). Indeed, the consequences of behavioral changes are particularly relevant with this technology. In contrast with other more passive media, the fact that in VR the participant is not a "spectator" but an "actor" raises the possibility that virtual agency could have consequences on real-world behavior (Brey, 1999, p. 8). Furthermore, the considerable cheapening and diffusion of VR technologies (including rudimentary cellphone-based hardware formats), with the consequent boom in domestic use, make this issue even more pressing. Virtual experiences—seemingly confined to the realm of the home—are becoming massive, so these presumably private practices could have ethically relevant public consequences.

In assessing the potential for VR to do good, a number of empirical and normative issues need to be considered. This is precisely the purpose of the overview I propose in this chapter. On the one hand, if we want to know what hopes and fears are justified, we must attend to the burgeoning interdisciplinary literature studying the impacts of VR on people's prosocial behaviors and moral capacities. Knowing which types of virtual experiences produce changes in attitudes, which reduce or increase implicit biases, or which factors of virtual exposure may cause lasting effects is a straightforward empirical question. On the other hand, whether these changes are desirable or undesirable is a normative problem. Even if there is agreement on the evidence generated on whether VR could have a particular influence, there may still be ethical divergences on whether the influence in question is good or bad. These empirical and normative levels will be intertwined in the subsequent discussion, showing that the assessment of the impact of VR on human morality is complex and nuanced.

---

social awareness have also been cataloged under the name of the "'co-presence for good' movement" (Nakamura, 2020, p. 56).

[2] Beyond computer-generated environments, 360° recorded videos that are projected on head-mounted displays can also be considered VR experiences, although these are generally less immersive than digitally designed environments (Slater and Sanchez-Vives, 2016, p. 35-37). Later I will show that some of the hopes and criticisms of using VR to improve socio-moral behavior have also been focused on these immersive videos, especially regarding empathy enhancement. Therefore, in this chapter, I will manage a broad conception of VR that can accommodate different developments and applications of virtually immersive technologies.

The structure of this chapter is as follows. I will begin by showing the rationale for using VR to improve human morality, pointing out the specific potential of this technology. Next, I will look at three cases where VR could notably influence morally significant abilities and prosocial types of behavior. I will then show that socio-moral projects through VR also have objectionable aspects by analyzing three specific problems. Finally, I will conclude with a series of reflections on the limitations of the chapter and with future lines of discussion.

2. THE RATIONALE FOR USING VR TO IMPROVE MORALITY

Human morality can be empirically studied and deliberately influenced through emerging technologies. On the one hand, knowledge about the foundations of morality in our species has considerably increased in recent decades. First, human morality has a neurobiological basis (Rueda, 2021). We know, for example, that certain brain lesions—either caused by accidents or by nontraumatic disorders—can affect our socio-moral capacities (Harlow, 1848; Damasio, 1994; Mendez, 2009). Neuroimaging studies also provide valuable information about the brain in action, showing which brain regions are more important in specific moral judgments or in morally problematic situations (Greene et al., 2002; Illes, 2003; Racine et al., 2005; Prehn & Heekeren, 2014). There is also ample evidence on the neurobiological basis of morality because it has been shown that neurochemical modulation and brain stimulation techniques can influence moral judgments and behaviors (Bourzac, 2016; Crockett, 2016; Levy et al., 2016; Di Nuzzo et al., 2018). Second, other branches have studied the evolutionary origin of morality. The building blocks of our moral psychology seem to have developed in the Pleistocene, when humans lived in small, interdependent, and close-knit communities (van Schaik et al., 2014; Burkart, 2018; Tomasello, 2018). The moral dispositions that remained ingrained in our forged evolutionary psychology hinder, according to some, the fulfillment of today's ethical aspirations to care for strangers who are distant in space and time (Persson & Savulescu, 2012). Third, we have further empirical evidence coming from "experimental ethics", that is, from studies (using methods mainly from cognitive sciences, empirical moral psychology, or behavioral economics) showing the psychological, cognitive, and behavioral mechanisms that influence our moral tendencies in both lay and expert persons (see Dworazik & Rusch, 2014; Aguiar et al., 2020).

On the other hand, emerging technologies can deliberately influence, for better or worse, our moral capacities. In particular, the field of study on "moral enhancement" has investigated how to use technological advances in beneficial ways to improve human morality. By 'moral enhancement', we generally mean using technological applications or science-based interventions to improve cognitive, emotional, motivational, and behavioral aspects related to morality (DeGrazia, 2014; Raus et al., 2014; Rueda 2020). Different methods could serve these purposes, such as, among others, emerging neurotechnologies (Earp et al., 2018; Di Nuzzo et al., 2018), various pharmaceuticals (Crockett, 2014; Levy et al., 2016; Lara, 2017), genetic technologies (Faust, 2008; Walker, 2009; Agar, 2010; Douglas & Devolder, 2013), artificial intelligence (Savulescu

& Maslen, 2015; Klincewicz, 2016; Giubilini & Savulescu, 2018; Lara & Deckers, 2020), and even VR itself (Rueda & Lara, 2020; Lara & Rueda, 2021).

In this chapter, I have an interest that goes beyond the use of VR for 'moral enhancement'. That term is contentious on both the 'enhancement' and 'moral' sides. To begin with, the term 'enhancement' is not without controversy (Gyngell & Selgelid, 2016). For instance, a predominant conception of 'enhancement' points to interventions that improve capabilities beyond what is normal (in a statistical sense) in a population (Daniels, 2000; Schwartz, 2005). In other words, an enhancement goes beyond what is typical in terms of concrete functionality in a given group. However, establishing what is 'normal' or 'typical' is often difficult (Gyngell & Selgelid, 2016; Rueda et al., 2021). We also risk confusing what is normal with what is desirable. Furthermore, going beyond the normal need may not always be a positive thing—increasing, for instance, certain abilities such as smell and hearing may be undesirable in many environments (Earp et al., 2014).

It is important, moreover, to reflect on what we mean by 'moral' in *moral* enhancement. It is difficult (or perhaps impossible) to offer a completely neutral (non-value-laden) description of the capabilities we regard as 'moral'. That is, it is challenging to articulate a definition of moral enhancement that does not implicitly have some normative positioning about what is desirable or undesirable in order to facilitate morality (DeGrazia 2014; Raus et al., 2014). Similarly, there are many rival conceptions of what is morally valuable in our societies. Some people may think that moral enhancements would help improve compassion, while others would think of interventions that would strengthen loyalty to their nation or devotion to their family (O'Neill et al., 2022). Given these difficulties, I will not restrict the discussion below to narrow views of morality or solely to strict uses of 'enhancement' in the extended (but controversial) sense mentioned above. Rather, I am interested in the varied impacts of VR on human morality in a broad sense.

Let us now consider the particularities of VR concerning its potential to affect the cognitive underpinnings of human morality. Although it has not been given the most systematic attention, VR is a very powerful technology for influencing the psychological and behavioral aspects of morality for several reasons. Immersive experiences in virtual environments can affect the plasticity of the human mind and modulate the contextual factors that shape personal conduct, even without being aware of the influence of these external factors (Madary & Metzinger, 2016). This is largely due to the feeling of 'presence', a main psychological characteristic of VR, by which we have the experience of "being there" (Heeter, 1992). In other words, the place illusion is crucial for having the sense of being in the virtual world (Slater, 2009). When the feeling of presence is properly attained, users may react in virtual environments in a manner analogous to how they would react in non-virtual ones (Slater et al., 2006; Felnhofer et al., 2015; Oh and Bailenson, 2017; Bailenson, 2018).

In addition to the immersive nature of virtual scenarios, it is worth mentioning the phenomenon of body ownership illusion. Head-mounted displays (HMDs) enable the simulation of virtual embodiment, namely, having the illusion that you are incorporated

in the body of an avatar in the first-person perspective. Interestingly, changing the virtual body changes the self, even to the point of having after-effects in subsequent real-life experiences (Maister et al., 2015; Slater et al., 2020). Indeed, the 'proteus effect' refers to the phenomenon of how altering our digital self-representation influences behavior, acting according to what may be expected from the identity of the virtual avatar—and whose effect sometimes extends beyond the virtual experience (Yee & Bailenson, 2007). For instance, being embodied in a superhero avatar when people are in danger increases helping behavior towards them (Rosenberg et al., 2013).

Another interesting feature of VR is its potential to produce an enormous range of possible experiences. Jeremy Bailenson labeled, in this sense, VR as an "experience generator" (Bailenson, 2018)—which reminds us of Robert Nozick's (1974) idea of the "experience machine". Others have spoken of VR as a powerful simulation medium (Ramirez & LaBarge, 2018). Both conceptions converge in the view that the fundamental element of VR is the type of experiences it allows us to live. We could have, albeit costly, many of these experiences in the non-virtual physical world. Others, in contrast, would be directly impossible. For example, to return to the previous mention, VR can (rudimentarily) simulate abilities such as the superpower of flight, so that we can have the experience of temporarily being a superhero. In view of its experience simulator character, it is not surprising that VR is increasingly associated with a kind of "experiential ethics" (Cotton, 2021, p. 31). A prominent appeal of VR for moral improvement is therefore its potential for "experiential moral learning". VR allows us to reinforce moral qualities through doing and establishing habits, which is fundamental to moral development according to ancient Greek thinkers and contemporary pragmatists (Rueda & Dore-Horgan, 2022). This possibility is especially seductive for subjects in limited social environments and who cannot develop actions that could improve their moral character, such as prisoners (Ligthart et al., 2022).

In light of this, VR can influence our morality by placing us in different environments that we feel are real, through body transfer to other avatars, and by providing us with morally relevant experiences. The following section will address some more concrete impacts of this technology.

3. DOMAINS OF SOCIO-MORAL IMPROVEMENT THROUGH VR

VR can influence human morality in many ways. In this section, I will focus on just a few domains. The following cases have been selected either because they have received considerable academic attention, because they have generated particular enthusiasm in civil society, or because they represent a mixture of both. I shall particularly focus on how VR technologies may affect empathic abilities, reduce implicit biases, and improve pro-environmental awareness and behavior.

3.1. The "empathy machine"

The use of VR with the intention of improving empathic skills has been one of the leading discussions. This issue is the subject of at least three fundamental controversies. The first is what we mean by empathy and what subtypes can be distinguished within it. The second is whether VR actually influences empathy. The third is if the role played by empathy, or any of its subtypes, is important for morality.

Let us start with the conceptual issue. The term 'empathy' comprises a set of diverse psychological capacities that enable us to feel or imagine the experiences of others. Although at least eight subtypes can be distinguished (Batson, 2009), it is common, in the literature about VR, to focus on two very different kinds of empathy (Fisher, 2017; Ramirez, 2017; Hamilton-Giachritsis et al., 2018; Francis et al., 2018; Seinfeld et al., 2018; van Loon et al., 2018; Bailenson, 2018; Schoeller et al., 2019). On the one hand, *emotional empathy* usually refers to the action of mirroring other people's emotions, mimicking their affective states almost as a reflex action. On the other hand, *cognitive empathy* refers to taking the other person's perspective, either by imagining the mental state of that subject from their point of view (imagine-other) or by imagining ourselves in that person's position (imagine-self) (Davis, 1980; 1983; Batson, 1997). To these two main subtypes, some also add a third *motivational* dimension to empathy (Zaki, 2014; Bailenson, 2018). Feeling or imagining the affective states of others may lead to alleviate their perceived suffering or to be proactively concerned about their well-being (Batson, 2015). This subtype, linked to altruistic motivation, is sometimes referred to as *empathic concern*.

After seeing the different ways of understanding empathy, let us now turn to the empirical question of how VR can affect these empathic skills. This is an important question because several studies have examined whether VR really works as an "empathy machine". The term 'empathy machine' was used by the filmmaker and visual storyteller Chris Milk (2015) to characterize VR during a TED talk, becoming a popular expression beyond academic circles. Some organizations, visual artists, and communicators with social concerns adopted this idea of using VR to empathize with, for example, refugees (Milk, 2015) or non-human animals.[3] But were these social projects effective in influencing empathy? To answer this question, I must first clarify the main technical characteristics of these experiences.

Most of the creators from civil society used videos recorded in 360º to be projected on the HMD. These cinematic experiences may give the impression of seeing the life of their protagonists in first person, but they do not allow us to embody interactive virtual avatars as such. Evidence in hand, these immersive videos or movies in VR glasses are

---

[3] See iAnimal360 of Animal Equality [Available at: https://ianimal360.com/; last access 31st January 2023] or the projects of People for the Ethical Treatment of Animals such as *I, Chicken*, *I, Orca*, and *Eye to Eye*.

not as powerful in eliciting empathy as their creators would have wished, nor significantly more advantageous in influencing empathic processes than other non-immersive media (Sundar et al., 2017; Archer & Finger, 2018; Bang & Yildirim, 2018; Weinel et al., 2018). The research of Schutte and Stilinović (2017) is one of the exceptions which showed a significant increase in empathic perspective-taking and empathic concern, in which participants did empathize more with the refugee protagonist of Milk's *Clouds Over Sidra* compared to control groups viewing it in a two-dimensional format. But, on balance, the review of the majority of evidence does not allow us to affirm that immersive 360º videos make VR a particularly effective empathy machine (Rueda & Lara, 2020; Sora-Domenjó, 2022).

Other lines of research are related to virtual embodiment. In these simulations, the aim is to increase empathy with the groups represented by the avatars in which the participants were incorporated. In general terms, this line has been more successful in eliciting prosocial motivation and more inclusive attitudinal changes than immersive video. As mentioned, virtual embodiment permits us to incorporate individuals in avatars that represent the social targets with whom we want to raise empathy (Rueda & Lara, 2020). In this way, changing our body depictions and social identities in VR may increase the tendency to take the perspective of the represented collective outside virtual environments. This strategy has been used to embody participants in avatars of other genders (Seinfeld et al., 2019), skin tone (Groom et al., 2009; Peck et al., 2013; Banakou et al., 2016; Hasler et al., 2017), ages (Oh et al., 2016; Hamilton-Giachritsis et al., 2018), members of disabled groups (Ahn et al., 2013; Chowdhury et al., 2019), homeless people (Herrera et al., 2018), and even non-animal species (Ahn et al., 2016).[4]

We should be cautious, however, in interpreting too optimistically the later results as clear examples of empathy enhancement. Here, unlike with 360º immersive videos, we are not discussing whether there have been significant influences of virtual embodiments, but rather what the causal mechanisms inducing those prosocial changes are. The theoretical disagreement is about whether the main effects are produced by mechanisms linked to empathy or not. Some studies have associated positive attitudinal changes after virtual embodiment with empathy, as in one study that showed how cognitive empathy was increased (van Loon et al., 2018), and, more importantly, in another longitudinal study in which impacts on perspective-taking ability were longer lasting (Herrera et al., 2018). Others have offered different cognitive explanations. At least regarding positive changes in implicit attitudes, some views challenge the idea that empathy-related mechanisms are the underlying force of these effects (Tsakiris, 2017; Bedder et al., 2019; Slater & Banakou, 2021). I shall attend to them in the following subsection when addressing the reduction of implicit biases.

Finally, I shall briefly consider the normative dimension of the controversy around VR and empathy enhancement. Is it desirable to improve empathy to behave more

---

[4] The term 'homuncular flexibility' refers to the high malleability of the body schema in VR, which facilitates even the virtual embodiment in avatars that do not have human appearance (Won et al., 2015a; 2015b).

morally? That question has sparked heated debates. On the one hand, for some authors, the role of empathy in morality is contested (Prinz, 2011a; 2011b; Bloom, 2016). On these views, not only is empathy not necessary for morality, but it sometimes hinders our aspirations for impartiality, as our empathic tendencies favor in-groups, i.e., those to whom we are already most motivated to help. Admittedly, we tend to have less problems in sharing the emotions of individuals with whom we share similarities, affinities, or familiarity (Bertrand et al., 2018). Therefore, according to this position, empathy may be counterproductive if we want to improve moral behavior in favor of out-groups.

On the other hand, empathy may play a positive contribution to moral behavior. This is the case of reflective and reason-guided conceptions of empathy (Persson & Savulescu, 2018; Rueda & Lara, 2020; Lara & Rueda, 2021), which claim that empathic perspective taking is sometimes a source of valuable moral reasons to care about the well-being of others. These views, however, may serve to criticize certain proposals that use VR for empathic purposes. For instance, emotive videos in an immersive 360º format are not ethically promising as long as they do not mainly foster perspective-taking capabilities (Rueda & Lara, 2020), which are less morally controversial compared to rudimentary forms of emotional contagion (see Prinz, 2011a; 2011b; Masto, 2015; Bloom, 2016). But these views, by contrast, could positively assess virtual embodiment initiatives as far as they make it easier to engage with the perspective of those with whom they have more difficulties, i.e., with the out-groups.[5] In this way, improving empathy can be beneficial (but of course not sufficient) for morality because of its epistemological functions (Coplan, 2011; Oxley, 2011; Masto, 2015) and its role in motivating behavior (Masto, 2015; Persson and Savulescu, 2018; Read, 2019).

To summarize, VR puts users in the shoes of individuals representing multiple collectives. As the evidence and normative positions are not entirely conclusive, the debate on the use of VR to influence empathy is likely to continue.

3.2. Reducing implicit biases

As shown in the case of empathy, VR may alter social cognition. In this subsection, I will shortly deal with the phenomena—somewhat related to the previous controversy—of implicit biases. Generally speaking, implicit biases are automatic stereotypes and unintentional prejudices that may influence our decisions and behavior, leading to harmful social injustices (Brownstein, 2019; Lin et al., 2020). So, an apparently attractive

---

[5] It should be noted, however, that being embodied in an avatar representing a particular group is not equivalent to having the same actual experience as a member of that group (Lara & Rueda, 2021). Moreover, in some cases, the risk of misrepresentation is high. In disability simulations, for example, the participant may feel the temporary deprivation of a capacity more acutely than the disabled person who is accustomed to his or her condition and who does not have the same negative perception. Although some disability simulations (not through VR) may increase empathy towards that group, they may also provide misleading information and increase discrimination in subtle ways (Silverman, 2015). I am grateful to Andrew T. Kissel for bringing this point to my attention.

option to influence our morality would be to reduce unconscious biases and implicit attitudes related to negative social stereotypes.

The process of virtual embodiment, again, is a promising way for achieving the reduction of implicit biases. By changing our bodily self-representation, VR can help decrease implicit biases regarding the collectives represented by the avatar in which users are incorporated. Tanvir I. Chowdhury and colleagues showed, for instance, that embodying an avatar in a wheelchair reduces the negative association towards disabled individuals and leads to a lower score on the Implicit Association Test (Chowdhury et al., 2019). They also revealed that the effects of simulating disabilities were significantly greater in VR compared to a non-immersive computer desktop.

Furthermore, racial embodiment has been another prominent case of study. Tabitha C. Peck and colleagues showed that embodying light-skinned participants in dark-skinned avatars reduced their implicit racial bias. Domna Banakou and coauthors showed, moreover, that the reduction of implicit racial bias may last even one week after only one virtual exposure (Banakou et al., 2016). Similarly, a study conducted by Béatrice S. Hasler et al. (2017) demonstrated that racial embodiment may also reverse in-group mimicry favoritism, independently of their level of implicit racial bias. It increases mimicry behavior with the virtual counterpart that shares the virtual race of the participants (but not their actual racial identity). However, there is one example of increasing implicit biases after racial embodiment. In fact, in the first experiment on full-body racial embodiment, conducted by Victoria Groom and colleagues, Caucasian participants increased their implicit racial biases after being embodied in dark-skinned virtual avatars (Groom et al., 2009). Subsequent research on racial embodiment has pointed out that a possible explanation for the above result is that Groom et al. placed participants in job interviews—arguably socially hostile, competitive scenarios in which self-representation is more fragile (Hasler et al., 2017; Bedder et al., 2019; Slater & Banakou, 2021).

That said, how are the changes in implicit biases induced by VR? Implicit biases, needless to say, have a complex nature with cognitive, social, and physical ingredients (Brownstein, 2019; Lin et al., 2020). Still, some explanations from the cognitive sciences are particularly appealing to try to make sense of these changes produced by virtual embodiment. Prominently, Rachel L. Bedder and colleagues proposed the mechanistic account of "bodily resonance" to explain previous cases of implicit bias reduction through VR (Bedder et al., 2019). This proposal is based on how we compared our self-image representations with other individuals. Our self-image representations include diverse characteristics such as group membership, physical, bodily, or aesthetic traits. Bodily resonance, then, refers to the cognitive mechanism of comparing those encoded features with other people to see which characteristics overlap and which diverge. This cognitive-based explanation converges with previous contributions about how changing our bodily representations may affect our self-image and subsequent social cognition of out-groups (Maister et al., 2015; Tsakiris, 2017; Farmer & Maister, 2017). Virtual embodiment fosters, therefore, new associations between our self-representation and others, which

may lead to a more positive appraisal of features that were not previously encoded in our self-image.

Let us now turn to the ethical terrain. At first glance, the possibility of reducing implicit biases seems ethically desirable. These implicit attitudes condition our decisions and conduct, even if we are not aware of them, and even if we would well-intentionally wish that they did not affect us. Therefore, many agents may wish to reduce implicit biases that reinforce problematic behaviors and stereotypes. We could consider these interventions as improvements in moral autonomy.

To this positive reading, however, two cautions should be added. First, as we have seen, the context and interactions that occur in virtual embodiment seem relevant in modulating the effects. Therefore, designers should avoid generating virtual experiences that produce a reactive self-identification that increases implicit biases towards disadvantaged groups. Indeed, avoiding undesired effects is important to avoid a backlash against the socio-moral uses of VR (Sora-Domenjó, 2022). Second, although VR can serve as a bias-reduction technology to help us mitigate problematic tendencies to which we are subtly predisposed, the use of this technology to eradicate social discrimination is limited. The case of racial embodiment is a clear example in this regard. Certainly, if we are predisposed to implicit racial biases, reducing them is a good thing. It even appears, according to neuroscientific research, that racial aversion has neural correlates—being particularly related to the activity of the amygdala (Gazzaniga, 2005; Douglas, 2008; 2013). However, implicit racial attitudes should not be simply confused with the phenomenon of racism. Racism is a complex social phenomenon with structural causes, often including conscious discrimination, and where noticeable power relations are at play. Therefore, strategies to combat racism—in addition to using initiatives such as racial embodiment to reduce implicit bias—should be complemented with further measures that tackle the social roots of the problem (Rueda & Lara, 2020).

3.3. Improving pro-environmental behavior

Environmental degradation and climate change are among the most pressing challenges of this century. Those large-scale problems are global in scope and have repercussions for future generations. To mitigate their most deleterious ravages, these ecological challenges require joint coordination along with reinforcing individual motivation (Rueda, 2020). The massive scale of environmental problems, unfortunately, can demotivate many individuals due to the limitations of our moral psychology that favors those close in space and time (Persson and Savulescu, 2012). The difficulties in switching behavior to tackle climate change, however, are not only a motivational problem but they may also respond to cognitive factors (Kulawska y Hauskeller, 2018, p. 377). Fortunately, as we shall see, VR technologies can help in raising awareness and encouraging pro-environmental behavior.

Mainly, VR leads to the transformation of current information-based environmental communication strategies into experience-based approaches (Plechatá et

al., preprint). Multiple studies have exploited the idea of using vivid virtual experiences to test the impacts of VR on pro-environmental behavior. For this purpose, influencing the *locus of control* of individuals is one promising way of increasing responsibility awareness for sustainable behavior. Environmental locus of control refers to having the internal perception that one's own behavior has a direct impact on the environment (Cleveland et al., 2005). In a seminal article, Sun Joo (Grace) Ahn and colleagues showed that the virtual experience of cutting a tree augmented the self-reported environmental locus of control and diminished paper consumption in the real world compared to print and video messages (Ahn et al., 2014).

There are more studies providing evidence on how showing the environmental impacts of personal decisions in VR may achieve behavioral changes in the non-virtual world. Participants receiving vivid messages about the energy used to heat and transport water during a virtual shower used cooler water during real-life hand washing than participants that were exposed to less vivid messages (Bailey et al., 2015). Furthermore, other strategies focus on narrowing the temporal perception of future impacts. Indeed, another way to influence environmental attitudes is accelerating the progress of time to see, for instance, how marine life is endangered by ocean acidification, which positively affects its connectedness with nature (Ahn et al. 2016). VR serves to provide, moreover, experiences intended to foster sustainable eating through the anticipation of the future bad impacts of our dietary carbon footprint. Adéla Plechatá and colleagues showed that VR was effective in changing to more environmentally friendly plant-based diets one week after the virtual exposure (Plechatá et al., 2022; Plechatá et al., preprint).

In short, influencing locus of control shows how one can affect the perception of one's own responsibility. Although moral responsibility is not exhausted in attributability, attributability is important for knowing how to delineate when an agent causally contributes to a phenomenon for which they may deserve praise or blame (Douglas, 2019). Thus, VR can narrow the attribution gap between an action and its environmental impact. This attribution gap can also be narrowed, moreover, by showing the future impacts of present actions. In this way, diverse phenomena of environmental degradation occurring in slow motion—such as marine acidification—can be perceived as events of greater proximity through its acceleration in VR. Therefore, the potential for promoting environmental conservation is not negligible (Millar, 2016).

So, if combating environmental challenges is normatively desirable, and given the beneficial impacts empirically shown in the above studies, VR can be an ethically valuable tool to improve our pro-environmental behaviors.

4. OBJECTIONS

Initiatives that seek to do good through VR are not exempt from criticism. In this section, I will elaborate on three risks of VR, showing how (seemingly desirable) socio-moral

projects can contentiously feed the complacency of the privileged, how some virtual experiences might develop undesirable moral traits or behaviors, and how there may be trade-offs for the well-being of users. While these objections have appeal, my purpose is not to argue that they are strong enough to disqualify all attempts to improve human morality through VR. Rather, what I intend to show is that these counterpoints may reduce part of the enthusiasm for using VR to improve imperfect human morality.

4.1. Swelling the complacency of the privileged

Doing good can make us feel good. Personal gratification is often an element of altruistic behavior—which may even lead us to ask whether there is such a thing as a totally disinterested altruistic action (May, 2011; Kraut, 2020). Long-standing philosophical controversies aside, this question is important in the case of VR projects that want to promote social good, with some particularities. Most VR users and content creators belong to high-income countries, which may give rise to some criticism.

Lisa Nakamura (2020) has particularly been very critical of projects that try to do social good originating from the VR industry. With video games and porn being the most popular applications of VR, marketing VR as a technology for empathy and justice can be a social washing strategy. According to Nakamura, several leaders of Big Tech and digital platform capitalism—including Mark Zuckerberg—have been keen to promote VR as a technology that fosters greater social connections and progress. Many immersive media practitioners and tech entrepreneurs have taken advantage of this "cultural alibi" to develop projects that offer virtual experiences from the perspective of people from disadvantaged groups (p. 49). However, this trend problematically leads to a form of "identity tourism" by privileged people who confuse their immersive experience with the reality of people from these groups (p. 54). In addition, these simulations are experiential practices that are limited in their political effectiveness, as they focus on individual experiences rather than on systemic changes that fight inequalities.

It is important to consider, moreover, how the viewpoints of underprivileged groups are portrayed. The types of simulations fostered by social VR projects can range, for example, from the experiences of refugees, people with disabilities, individuals in prison or solitary confinement, the homeless, factory-farmed animals, to the perspective of people of other different races or genders. Thus, the issue of ethical representation is crucial (Brey, 1999). It is worth asking who represents whom, especially when it comes to creating experiences that want to bring the standpoint of marginalized collectives or stigmatized identities to the fore. A key recommendation is then to involve the underprivileged groups that are represented in the development of these virtual simulations (Rueda & Lara, 2020). Such participation would help to mitigate concerns that these simulations produce misrepresentations and inculcate false beliefs, especially when they are intended to convey the first-person experiences of members of intersectional out-groups (see Ramirez et al., 2021).

Ultimately, this objection reminds us of the importance of the power relations that underlie the different parties involved in the socio-moral projects of VR. This factor is relevant, certainly, when the production and consumption of seemingly laudable projects that seek to combat various social injustices are carried out by privileged collectives.

4.2. The real perils of bad virtual actions

For some people, the possibility of being able to abandon everyday morals may be an attractive feature of VR. Assaulting, stealing, killing, or raping are actions available in some virtual environments. For others, the above virtual experiences are examples of unacceptably risky content. In any case, the possibility of virtually performing courses of action that would generally be condemnable in non-virtual environments raises at least two interesting ethical questions.

On the one hand, the ethical status of virtual actions is a question to be asked (Brey, 1999). Is it wrong to treat virtual avatars badly? Similarly, which virtual behavior with morally reprehensible content—such as murder, rape, or pedophilia—should not be allowed? This has led to interesting ethical discussions, for example, on how to justify violent video games in which killing is a common practice that generates little scandal, as opposed to other behaviors that cause greater rejection, such as virtual pedophilia— the so-called "gamer's dilemma" (Luck, 2009).[6]

On the other hand, we may wonder whether bad virtual actions may lead to bad real-world actions or even to the development of an evil moral character. This is the issue that interests us the most here. Consider the case of violent VR games. Does virtual violence increase violent behavior offline? This is a typical concern in computer and gaming ethics. If violent VR games desensitize individuals to certain types of aggressive behavior, violent conduct by these persons may increase in the non-virtual world. Of course, this is an empirical question on which there is no clear evidence (Spiegel, 2018, p. 1542). However, it is often argued that this possibility may be harmful in terms of moral development, relying on an analogy with a typical Kantian argument. According to Philip Brey (1999, p. 9), "if disrespectful treatment of animals causes disrespectful treatment of human beings, then disrespectful treatment of virtual characters, which may be even more similar to such treatment of real humans, will have the same consequence". Furthermore, this could be more problematic with more advanced VR, where we may encounter more realistic gaming characters with high-fidelity in human-like appearances (Kade, 2016, p. 82; Slater et al., 2020).

As it can be seen, this objection depends on an empirical premise, for which we do not yet have solid evidence. If it were shown that objectionable virtual behaviors do not translate into an increase in objectionable offline behaviors, this objection would not hold. Yet, the uncertainty does not allow us to dismiss this concern entirely. The long-

---

[6] For ethical arguments that would allow virtual pedophilia through computer-generated child pornography (and also child-like sex robots), see Moen & Sterri (2018).

term development of undesirable moral traits is a relevant worry as regards VR (Ramirez & LaBarge, 2020). The absence of long-term longitudinal studies on the effects of VR exposure (Madary & Metzinger, 2016; Slater, 2021) is an added motivation for this type of research—which may resolve the issue of this objection's validity in the future.

4.3. Concerns about the well-being of VR users

Another factor that would diminish enthusiasm of employing VR for moral improvement would arise if virtual experiences reduce the users' well-being. There are various ways in which VR could go against users' interests. Broadly speaking, virtual experiences may cause psychological and health problems for users. This risk is especially prominent in cases in which extended immersion may lead to neglecting the physical environment, fostering depersonalization—i.e., seeing one's own physical body as strange—, losing interest in the non-virtual world, increasing long-term social isolation, and even leading to bodily and child neglect (Madary & Metzinger, 2016; Spiegel, 2018; Slater et al., 2020). Moreover, in addition to being potentially addictive, the use of VR is sometimes accompanied by other addictive behaviors—especially the consumption of psychotropic drugs during gaming (Lorenz, 2020).

All in all, although we need more evidence on the after-effects of virtual simulations and the lasting consequences of long-term exposure, there is the risk of creating bodily and psychological harm. This is even more problematic when VR exploits the vulnerabilities of consumers such as teenagers, people with addictive tendencies, or individuals with mental disorders.

Another important interest of VR users has to do with privacy and personal data generated by their interactions with this technology (O'Brolcháin et al., 2016; Madary & Metzinger, 2016; Slater et al., 2020). Many VR applications are produced by platforms that may have a commercial interest in collecting user data. Therefore, the correct protection of personal data is key in order not to unduly threaten the privacy of users, especially when carrying out more sensitive behaviors in the virtual world. This concern is even more important if the use of VR systems that incorporate eye-tracking devices or emotion-capture technologies becomes widespread (O'Brolcháin et al., 2016).

Furthermore, there are issues specifically related to the use of VR in morally salient scenarios. This problem arises, in particular, when VR places users in immersive situations that stand out as morally demanding. There is a very interesting example from psychological research on moral behavior in virtual environments. Kathryn Francis and colleagues (2016; 2017; 2018) simulated the footbridge version of the trolley dilemma (Foot, 1967), in which the only way to save the five innocent people trapped on the rails from a runaway trolley is by pushing a large person to die from a bridge (Thomson, 1985). Whereas previous research had already simulated the lever version of the trolley dilemma (Navarrete et al., 2012; Patil et al., 2014; Skulmowski et al., 2014)—and even trolley-like scenarios reproducing traffic dilemmas with autonomous cars (Sütfeld et al., 2017; Faulhaber et al., 2019)—, Erick Ramirez and Scott LaBarge (2020) found the footbridge

version experiments especially controversial. In their view, the vividness of the later sacrificial moral dilemma may have generated stress and even trauma in the participants. Although elsewhere I have tried to downplay these concerns (Rueda, 2021), those risks should not be completely underestimated in the future.

According to Ramirez it would be unethical in research contexts to place participants in experiences that we would consider unacceptable in the non-virtual world, what he calls "the principle of equivalence" (Ramirez, 2019). That idea is interesting beyond research and scientific domains, though. Would it be problematic to place users in stressful situations but from which we could obtain a socio-moral benefit? Would it be ethically impermissible? To answer this last question, the magnitude of the risks and benefits should first be sized.

Causing lasting psychological damage would be arguably unacceptable if the induced moral changes are minor, short-lived, or highly contextual. But if, on the contrary, the psychological costs are small and the benefits are high in terms of attitudinal and behavioral changes, we could be dealing with permissible simulations. Consider, for instance, the model proposed by Mel Slater and Domna Banakou (2021) about the Golden Rule Embodiment Paradigm for promoting helping behavior and diminishing antisocial behavior. This paradigm for using virtual embodiment to promote prosocial behavior is described, according to those authors, as follows:

> First, participants must be complicit in an action that causes harm to another person. Second, later, they must reexperience that episode from the embodied viewpoint of the victim, being able to observe their own previous actions (or acquiescence) during the course of the harmful event from that viewpoint. (Slater and Banakou, 2021, p. 506)

These kinds of virtual experiences help us to put ourselves in the position of others, to avoid treating them in a way that we would not want to be treated, or to treat them like we would like to be treated—as the famous golden rule in moral philosophy recommends. To be sure, experiencing the victim's point of view, and perhaps also being the perpetrator, can be unpleasant for users. But if these interventions are effective in improving prosocial behavior, that distress may be a minor cost justifiable by the resulting major benefits. As mentioned in the second section, VR (as much as it may create transient discomfort) can help in experiential moral learning, by training us in the development of moral character. To borrow a popular saying, in morality, as in life, a smooth sea never made a good sailor.

5. CONCLUDING THOUGHTS

Doing good with VR is a respectable aspiration, but not without its difficulties. Virtual environments and virtual embodiment generate opportunities for simulating experiences that help improve our moral tendencies. In this chapter, I have shown three areas where

there may be particular—albeit sometimes contested—potential. VR can help foster empathic skills, reduce implicit biases, and enhance pro-environmental behaviors.

However, enthusiasm for using VR to improve human morality should be restrained. Many of the socio-moral projects may be comforting to privileged individuals and collectives, but the extent to which they improve the lives of the most disadvantaged must be studied. It is also worth considering the negative impacts that morally problematic virtual actions and risky content may have on the development of moral character. Finally, the well-being of users must be protected, trying to minimize the adverse effects that virtual experiences may have on their mental and bodily health, in addition to safeguarding the privacy of sensitive personal data.

I hope that the analysis offered in this chapter will encourage future research to fill the alarming gap concerning VR's long-term effects. As shown, empirical evidence is crucial for weighing the (beneficial or detrimental) impacts of VR from an ethical perspective. Consequently, not only should moral philosophers and applied ethicists be heartened to enter this debate, but scientists should also be encouraged to generate the strongest possible evidence. Whether VR becomes an ally for social good is an exciting topic for which there are still many unresolved questions.

REFERENCES

Agar, N. (2010). Enhancing genetic virtue? *Politics and the Life Sciences*, *29*(1), 73-75.

Aguiar, F., Gaitán, A. & Viciana, H. (2020). *Una introducción a la ética experimental* [An Introduction to Experimental Ethics]. Madrid: Cátedra.

Ahn, S. J., Le, A. M. T., & Bailenson, J. (2013). The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychology*, *16*(1), 7-38.

Ahn, S. J., Bailenson, J. N., & Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behavior*, 39, 235–245. 10.1016/j.chb.2014.07.025

Ahn, S. J. G., Bostick, J., Ogle, E., Nowak, K. L., McGillicuddy, K. T., & 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. https://doi.org/10.1111/jcc4.12173.

Archer, D., & Finger, K. (2018). Walking in another's virtual shoes: Do 360-degree video news stories generate empathy in viewers? New York, NY: Tow Center for Digital Journalism, Columbia University. doi: 10.7916/D8669W5C

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. New York: WW Norton & Company.

Bailey, J. O., Bailenson, J. N., Flora, J., Armel, K. C., Voelker, D., & Reeves, B. (2015). The impact of vivid messages on reducing energy consumption related to hot water use. *Environment and Behavior*, 47(5), 570-592.

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 601.

Bang, E., & Yildirim, C. (2018). Virtually empathetic?: Examining the effects of virtual reality storytelling on empathy. In *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I 10* (pp. 290-298). Springer International Publishing.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin*, *23*(7), 751-758.

Batson, C. D. (2009). These things called empathy: Eight related but distinct phenomena. In. J. Decety and W. Ickes (eds). *The Social Neuroscience of Empathy* (pp. 3–16), Cambridge, MA: MIT Press.

Batson, C. D., Lishner, D. A., and Stocks, E. L. (2015). The empathy-altruism hypothesis. In D. A. Schroeder and W. G. Graziano (eds.), *The Oxford Handbook of Prosocial Behavior* (pp. 259–281), Oxford: Oxford University Press.

Bedder, R. L., Bush, D., Banakou, D., Peck, T., Slater, M., & Burgess, N. (2019). A mechanistic account of bodily resonance and implicit bias. *Cognition*, *184*, 1-10.

Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, 26.

Bloom, P. (2016). *Against Empathy*. London: Bodley Head.

Bourzac, K. (2016). Neurostimulation: bright sparks. *Nature*, *531*(7592), S6-S8.

Brey, P. (1999). The ethics of representation and action in virtual reality. *Ethics and Information Technology*, *1*(1), 5–14. https://doi.org/10.1023/A:1010069907461

Brownstein, M. (2019). Implicit bias. In E. N. Zalta (Ed). *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition); Stanford University: Stanford, CA, USA. Available at: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=implicit-bias [last access 2nd February 2023].

Burkart, J. M., Brügger, R. K., & Van Schaik, C. P. (2018). Evolutionary origins of morality: Insights from non-human primates. *Frontiers in Sociology*, *3*, 17.

Chowdhury, T. I., Ferdous, S. M. S., & Quarles, J. (2019). VR disability simulation reduces implicit bias towards persons with disabilities. *IEEE transactions on visualization and computer graphics*, *27*(6), 3079-3090.

Cleveland, M., Kalamas, M., & Laroche, M. (2005). Shades of green: linking environmental locus of control and pro-environmental behaviors. *Journal of consumer marketing*, *22*(4), 198-212.

Coplan, A. (2011). Understanding empathy: its features and effects. In P. Goldie and A. Coplan (eds.). *Empathy: Philosophical and Psychological Perspectives* (pp. 3–18), New York, NY: Oxford University Press.

Cotton, M. (2021). The Ethical Dimensions of Virtual Reality. In *Virtual Reality, Empathy and Ethics* (chapter 2, pp. 23-41). Cham: Springer. https://doi.org/10.1007/978-3-030-72907-3

Crockett, M. J. (2014). Moral bioenhancement: a neuroscientific perspective. *Journal of medical ethics*, *40*(6), 370-371.

Crockett, M.J. (2016). Morphing morals: Neurochemical modulation of moral judgment and behavior. In Liao MS, ed. *Moral Brains: The Neuroscience of Morality* (pp. 237–45). New York: OUP.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam.

Daniels, N. (2000). Normal functioning and the treatment-enhancement distinction. *Cambridge Quarterly of Healthcare Ethics*, *9*(3), 309–322. https://doi.org/10.1017/s0963180100903037

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catal. Select. Doc. Psychol*. 10:85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

DeGrazia, D. (2014). Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of medical ethics*, *40*(6), 361-368.

Douglas, T. (2008). Moral enhancement. *Journal of applied philosophy*, *25*(3), 228-245.

Douglas, T. (2013). Moral enhancement via direct emotion modulation: a reply to John Harris. *Bioethics*, *27*(3), 160-168.

Douglas, T. (2019). Enhancement and desert. *Politics, Philosophy & Economics*, *18*(1), 3-22.

Di Nuzzo, C., Ferrucci, R., Gianoli, E., Reitano, M., Tedino, D., Ruggiero, F., & Priori, A. (2018). How brain stimulation techniques can affect moral and social behaviour. *Journal of cognitive enhancement*, *2*, 335-347.

Douglas, T., & Devolder, K. (2013). Procreative altruism: beyond individualism in reproductive selection. *Journal of Medicine and Philosophy*, *38*(4), 400-419.

Dworazik, N. & Rusch, H. (2014). A Brief History of Experimental Ethics. In C. Luetge, H. Rusch & M. Uhl (Eds.), *Experimental Ethics. Toward an Empirical Moral Philosophy* (pp. 38–56). Cham: Springer.

Earp, B.D., Douglas, T., Savulescu J. Moral neuroenhancement. In L.S.M. Johnson & K.S. Rommelfanger (eds). *The Routledge Handbook of Neuroethics* (pp. 166–84). New York: Routledge; 2018.

Farmer, H., & Maister, L. (2017). Putting Ourselves in Another's Skin: Using the Plasticity of Self-Perception to Enhance Empathy and Decrease Prejudice. In *Social Justice Research*, 30(4), 323–354. https://doi.org/10.1007/s11211-017-0294-1

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., ... & König, P. (2019). Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. *Science and engineering ethics*, *25*(2), 399-418.

Faust, H. S. (2008). Should we select for genetic moral enhancement? A thought experiment using the MoralKinder (MK+) haplotype. *Theoretical medicine and bioethics*, *29*, 397-416.

Felnhofer, A., Kothgassner, O. D., Schmidt, M., Heinzle, A. K., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2015). Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies*, *82*, 48-56.

Fisher, J. A. (2017). Empathic actualities: toward a taxonomy of empathy in virtual reality. In N. Nunes, I. Oakley, and V. Nisi (eds.). *Interactive Storytelling. ICIDS 2017. Lecture Notes in Computer Science* (pp. 233–44), vol. 10690, Cham: Springer.

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*, *11*(10), e0164374.

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports*, *7*(1), 1-11.

Francis, K. B., Gummerum, M., Ganis, G., Howard, I. S., & Terbeck, S. (2018). Virtual morality in the helping professions: Simulated action and resilience. *British Journal of Psychology*, *109*(3), 442-465.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 5, 5–15.

Gazzaniga, M. S. (2005). *The Ethical Brain*. New York, NY: Dana Press.

Giubilini, A., & Savulescu, J. (2018). The artificial moral advisor. The "ideal observer" meets artificial intelligence. *Philosophy & technology*, *31*(2), 169-188.

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.

Groom, V., Bailenson, J. N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, *4*(3), 231-248.

Gyngell, C. & Selgelid, M. J. (2016). Human Enhancement: Conceptual Clarity and Moral Significance. In Steve Clarke, Julian Savulescu, C. A. J. Coady, Alberto Giubilini, and Sagar Sanyal (Eds.), *The Ethics of Human Enhancement: Understanding the Debate*. Oxford: OUP.

Hamilton-Giachritsis, C., Banakou, D., Garcia Quiroga, M., Giachritsis, C., & Slater, M. (2018). Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Scientific reports*, *8*(1), 1-10.

Harlow, J. M. (1848). Passage of an iron rod through the head. *The Boston Medical and Surgical Journal*, *39*(20), 389–93.

Hasler, B. S., Spanlang, B., & Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PloS one*, *12*(4), e0174965.

Heater, C. (1992). Being there: The subjective experience of presence. *Presence*, *1*(2), 262-271.

Illes, J. (2003). Neuroethics in a new era of neuroimaging. *AJNR: American Journal of Neuroradiology*, *24*(9), 1739.

Kade, D. (2016). Ethics of virtual reality applications in computer game production. *Philosophies*, *1*(1), 73–86. https://doi.org/10.3390/philosophies1010073

Klincewicz, M. (2016). Artificial intelligence as a means to moral enhancement. Studies in Logic, Grammar and Rhetoric, 48(1), 171–187. https://doi.org/10.1515/slgr2016-0061.

Kraut, R. (2020). Altruism. In E. N. Zalta (Ed). *The Stanford Encyclopedia of Philosophy* 2016 (Fall 2020 Edition); Stanford University: Stanford, CA, USA. Available at: https://plato.stanford.edu/entries/altruism/ [last access 31st January 2023].

Kulawska, A., & Hauskeller, M. (2018). Moral Enhancement and Climate Change: Might it Work?. *Royal Institute of Philosophy Supplements*, 83, 371-388.

Lara, F. (2017). Oxytocin, empathy and human enhancement. *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia*, *32*(3), 367-384.

Lara, F., & Deckers, J. (2020). Artificial intelligence as a socratic assistant for moral enhancement. *Neuroethics*, *13*(3), 275-287.

Lara, F. & Rueda, J. (2021). Virtual Reality not for "being someone" but for "being in someone else's shoes": Avoiding misconceptions in empathy enhancement. *Frontiers in Psychology*, 12: 3674. https://doi.org/10.3389/fpsyg.2021.741516

Levy, N., Douglas, T., Kahane, G., Terbeck, S., Cowen, P. J., Hewstone, M., & Savulescu, J. (2014). Are you morally modified? The moral effects of widely used pharmaceuticals. *Philosophy, psychiatry, & psychology*, *21*(2), 111.

Ligthart, S., Meynen, G., Biller-Andorno, N., Kooijmans, T., & Kellmeyer, P. (2022). Is virtually everything possible? The relevance of ethics and human rights for introducing extended reality in forensic psychiatry. *AJOB neuroscience*, *13*(3), 144-157.

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*(1), 31-36.

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 Robotics AI*, *3*(FEB), 1–23. https://doi.org/10.3389/frobt.2016.00003

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends in cognitive sciences*, *19*(1), 6-12.

Masto, M. (2015). Empathy and its role in morality. *The Southern Journal of Philosophy*, *53*(1), 74-96.

May, J. (2011). Psychological Egoism. *Internet Encyclopedia of Philosophy*. Available at: https://iep.utm.edu/psychological-egoism/ [last access 31st January 2023].

Mendez, M. F. (2009). The neurobiology of moral behavior: Review and neuropsychiatric implications. *CNS spectrums*, *14*(11), 608-620.

Millar, H. (2016). Can Virtual Reality Emerge As a Tool for Conservation? *Yale Environment 360*. 26th June 2016. Available at: https://e360.yale.edu/features/can_virtual_reality_emerge_as_a_tool_for_conservation [last access 28th January 2023].

Milk, C. (2015). How Virtual Reality Can Create the Ultimate Empathy Machine. *TED Talk*. Available online at: https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine [last access 31st January 2023].

Moen, O. M., & Sterri, A. B. (2018). Pedophilia and computer-generated child pornography. In *The Palgrave handbook of philosophy and public policy* (pp. 369-381). Cham: Palgrave Macmillan.

Nakamura, L. (2020). Feeling good about feeling bad: Virtuous virtual reality and the automation of racial empathy. *Journal of Visual Culture*, 19(1), 47–64. https://doi .org/10.1177/1470412920906259

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

O'Brolcháin, F., Jacquemard, T., Monaghan, D., O'Connor, N., Novitzky, P., & Gordijn, B. (2016). The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy. *Science and Engineering Ethics*, *22*(1), 1–29. https://doi.org/10.1007/s11948-014-9621-1.

Oh, S. Y., and Bailenson, J. (2017). Virtual and augmented reality. In P. Rössler, C. A. Hoffner and L. Zoonen (eds.). *The International Encyclopedia of Media Effects* (pp. 1–16). John Wiley & Sons.

Oh, S. Y., Bailenson, J., Weisz, E., & Zaki, J. (2016). Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, *60*, 398-410.

O'Neill, E., Klincewicz, M., & Kemmer, M. (2022). Ethical Issues with Artificial Ethics Assistants. In C. Véliz (Ed.), *The Oxford Handbook of Digital Ethics*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198857815.013.17

Oxley, J. C. (2011). *The Moral Dimensions of Empathy. Limits and Application in Ethical Theory and Practice*. New York, NY: Palgrave Macmillan.

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.

Plechatá, A., Morton, T., Perez-Cueto, F. J., & Makransky, G. (2022). A randomized trial testing the effectiveness of virtual reality as a tool for pro-environmental dietary change. *Scientific Reports*, *12*(1), 14315.

Plechatá, A., Hielkema, M., Merkl, L. M., Makransky, G., & Frøst, M. B. (2022). Fast Forward: Influencing the future in virtual reality increases real-life pro-environmental behaviour. Preprint available at https://assets.researchsquare.com/files/rs-2338078/v2/67fd5603-97eb-4a57-ba34-b5d1d4eb3969.pdf?c=1671652836 [last access 28th January 2023].

Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, *22*(3), 779-787.

Persson, I., & Savulescu, J. (2012). *Unfit for the future: The need for moral enhancement*. Oxford: Oxford University Press.

Persson, I., & Savulescu, J. (2018). The moral importance of reflective empathy. *Neuroethics*, *11*(2), 183-193.

Prehn K, & Heekeren H. (2014). Moral brains—Possibilities and limits of the neuroscience of ethics. In Christen M, van Schaik C, Fischer J, Huppenbauer M, Tanner C, eds. *Empirically Informed Ethics: Morality between Facts and Norms* (pp. 137–57). Cham: Springer.

Prinz, J. (2011a). Against empathy. *The Southern Journal of Philosophy*, *49*, 214-233.

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

Racine, E., Bar-Ilan, O., & Illes, J. (2005). fMRI in the public eye. *Nature Reviews Neuroscience*, *6*(2), 159-164.

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

Ramirez, E. J. (2019). Ecological and ethical issues in virtual reality research: A call for increased scrutiny. *Philosophical Psychology*, *32*(2), 211–233. https://doi.org/10.1080/09515089.2018.1532073

Ramirez, E. J., & LaBarge, S. (2020). Ethical issues with simulating the Bridge problem in VR. *Science and Engineering Ethics*, *26*(6), 3313-3331.

Ramirez, E. J., Elliott, M., & Milam, P. E. (2021). What it's like to be a _: why it's (often) unethical to use VR as an empathy nudging tool. *Ethics and Information Technology*, *23*(3), 527-542.

Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterckx, S. (2014). On Defining Moral Enhancement: A Clarificatory Taxonomy. *Neuroethics*, *7*(3), 263–273. https://doi.org/10.1007/s12152-014-9205-4

Read, H. (2019). A typology of empathy and its many moral forms. *Philosophy Compass*, *14*(10), e12623.

Rosenberg, R. S., Baughman, S. L., & Bailenson, J. N. (2013). Virtual Superheroes: Using Superpowers in Virtual Reality to Encourage Prosocial Behavior. *PLoS ONE*, *8*(1), 1–9. https://doi.org/10.1371/journal.pone.0055003

Rueda, J. (2020). Climate Change, Moral Bioenhancement and the Ultimate Mostropic. *Ramon Llull Journal of Applied Ethics*, 11, 277-303.

Rueda, J. (2021). Socrates in the fMRI Scanner: The Neurofoundations of Morality and the Challenge to Ethics. *Cambridge Quarterly of Healthcare Ethics*, 30 (4), 606-612. https://doi.org/10.1017/S0963180121000074

Rueda, J. (2022). Hit by the virtual trolley: When is experimental ethics unethical? *Teorema*, 41(1), 7-27.

Rueda, J. & Lara, F. (2020). Virtual Reality and Empathy Enhancement: Ethical Aspects. *Frontiers in Robotics and AI*, 7: 506984. https://doi.org/10.3389/frobt.2020.506984

Rueda, J., García-Barranquero, P. & Lara, F. (2021). Doctor, Please Make Me Freer: Capabilities Enhancement as a Goal of Medicine. *Medicine, Health Care & Philosophy* 24, 409-419. https://doi.org/10.1007/s11019-021-10016-5

Rueda, J. & Dore-Horgan, E. (2022). A virtual prosthesis for morality? Experiential learning through XR technologies for autonomy enhancement of psychiatric offenders. *AJOB Neuroscience*, 13 (3), 163-165. https://doi.org/10.1080/21507740.2022.2082584

Schoeller, F., Bertrand, P., Gerry, L. J., Jain, A., Horowitz, A. H., & Zenasni, F. (2019). Combining virtual reality and biofeedback to foster empathic abilities in humans. *Frontiers in psychology*, *9*, 2741.

Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and emotion*, *41*, 708-712.

Schwartz, P. H. (2005). Defending the distinction between treatment and enhancement. *The American Journal of Bioethics*, *5*(3), 17-19.

Savulescu, J., y Maslen, H. (2015). Moral enhancement and artificial intelligence: Moral AI? En J. Romportl, E. Zackova, y J. Kelemen (eds.), *Beyond artificial intelligence. The disappearing human-machine divide* (pp. 79–95). Springer.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D., ... & Sanchez-Vives, M. V. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Scientific reports*, *8*(1), 2692.

Silverman, A. M. (2015). The perils of playing blind: Problems with blindness simulation and a better way to teach about blindness. *Journal of Blindness Innovation and Research*, 5 (2). https://nfb.org/images/nfb/publications/jbir/jbir15/jbir050201.html

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS one*, *1*(1), e39.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549-3557.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74.

Slater, M. (2021). Beyond Speculation About the Ethics of Virtual Reality: The Need for Empirical Results. *Frontiers in Virtual Reality*, *2*(August), 1–4. https://doi.org/10.3389/frvir.2021.687609

Slater, M., & Banakou, D. (2021). The Golden Rule as a Paradigm for Fostering Prosocial Behavior With Virtual Reality. *Current Directions in Psychological Science*, *30*(6), 503–509. https://doi.org/10.1177/09637214211046954

Spiegel, J. S. (2018). The Ethics of Virtual Reality Technology: Social Hazards and Public Policy Recommendations. *Science and Engineering Ethics*, *24*(5), 1537–1550. https://doi.org/10.1007/s11948-017-9979-y

Sora-Domenjó, C. (2022). Disrupting the "empathy machine": The power and perils of virtual reality in addressing social issues. *Frontiers in Psychology*, *13*(September), 1–13. https://doi.org/10.3389/fpsyg.2022.814565

Sundar, S. S., Kang, J., & Oprean, D. (2017). Being there in the midst of the story: How immersive journalism affects our perceptions and cognitions. *Cyberpsychology, behavior, and social networking*, *20*(11), 672-682.

Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: applicability of value-of-life-based models and influences of time pressure. *Frontiers in behavioral neuroscience*, *11*, 122.

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

Tomasello, M. (2018). Precís of a natural history of human morality. *Philosophical Psychology*, *31*(5), 661-668.

Tsakiris, M. (2017). The multisensory basis of the self: from body to identity to others. *Quarterly journal of experimental psychology*, *70*(4), 597-609.

van Schaik, C., Burkart, J.M., Jaeggi, A.V., & von Rohr, C.R. (2014) Morality as a biological adaptation—An evolutionary model based on the lifestyle of human foragers. In Christen M, van Schaik C, Fischer J, Huppenbauer M, Tanner C, eds. *Empirically Informed Ethics: Morality between Facts and Norms* (pp. 65–84). Cham: Springer; 2014.

Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS one*, *13*(8), e0202442.

Walker, M. (2009). Enhancing genetic virtue: A project for twenty-first century humanity?. *Politics and the Life Sciences*, *28*(2), 27-47.

Won, A. S., Bailenson, J., Lee, J., & Lanier, J. (2015a). Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, *20*(3), 241-259.

Won, A. S., Bailenson, J., and Lanier, J. (2015b). Homuncular flexibility: the human ability to inhabit nonhuman avatars. In R. A. Scott, S. M. Kosslyn and M. Buchmann (eds.) *Emerging Trends in the Social and Behavioral Science: An Interdisciplinary Searchable, and Linkable Resources* (pp. 1–16), Hoboken, NJ: John Wiley & Sons.

Weinel, J., Cunningham, S., & Pickles, J. (2018). Deep subjectivity and empathy in virtual reality: a case study on the autism TMI virtual reality experience. *New Directions in Third Wave Human-Computer Interaction: Volume 1-Technologies*, 183-203.

Yee, N., and Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, *33*(3), 271-290.

Zaki, J. (2014). Empathy: a motivated account. *Psychological bulletin*, *140*(6), 1608.