*Routledge Studies in Contemporary Philosophy*

# EXPLORING EXTENDED REALITIES

## METAPHYSICAL, PSYCHOLOGICAL, AND ETHICAL CHALLENGES

Edited by
Andrew Kissel and Erick José Ramirez

# 9 Moral Behavior in Virtual Reality

*Eugy Han and Jeremy N. Bailenson*

Research areas surrounding morality evolve over time, given that social, cultural, and political contexts under which these behaviors arize matter. And as societies evolve and new landscapes emerge, it is critical to ask these questions within these contexts. One such context to consider is the virtual world, more specifically social worlds accessed through virtual reality (VR) headsets. Much like how internet communities cultivate their own norms and cultures, it can be expected that the same will apply to worlds within VR. It is especially important that we consider what these norms and cultures look like within social VR given the increasing popularity and interest in the "metaverse,"[1] as well as the unique affordances that VR provides.

The metaverse was first introduced as a fictitious concept in science fiction novels, such as Neal Stephenson's *Snow Crash* (1992), in which the term was first coined. The metaverse can be best described as a virtual world pieced together by digital landscapes and inhabited by digital beings. However, this idea of technology enhancing who and where we are has been explored by writers before then, as seen in novels such as William Gibson's *Neuromancer* (1984), Vernor Vinge's *True Names* (1981), or Philip K. Dick's *The Days of Perky Pat* (1990). What such seemingly distant but increasingly reasonable societies share in common is the need to reexamine how people's behaviors change as core concepts in life – such as the body, our memories, and our personalities – are transformed.

And while it is certainly not reasonable or productive to induce fear of our society heading in the direction of the worlds depicted in such works of fiction, it should be noted that some of the world's most influential shapers of technology are heavily investing their resources and attention to developing their own metaverses. Consequently, there is a growing need to understand how people's moral behaviors change in such environments. As the technology does not exist yet to support the fully realized version of the metaverse, how it is built can be informed by what we can understand through research and history.

DOI: 10.4324/9781003359494-13

A common misconception of the metaverse is that it is intended to be a somewhat veridical representation of the physical world. Experiences within the metaverse are frequently dismissed as not being photographically or behaviorally real enough. However, it is far from veridical. Research on how the affordances of virtual environments, such as what you look like, where you are, or who you are with, affect the way you perceive yourself and behave. We can already see examples of new practices, norms, and interactions taking place – ones that have been shaped by the affordances of different social VR worlds. One such example comes from the spatial property of VR. Many social VR platforms can track users' full-body movement and allow for six degrees of freedom to move around physically. This physicality is novel compared to what can be done in traditional social media or 2D virtual worlds. This affordance has bred novel forms of practices, such as users running around, shouting, and blocking others' views as a way to harass others (Freeman et al., 2022). Another example of a norm found within social VR is what Zheng and colleagues (2023) describe as *immersive dwellers* (or *mirror dwellers*), who are individuals in expensive avatars that spend hours staring at themselves in front of mirrors in public spaces. Some dwellers even sleep inside social VR or leave their headsets while their avatars are "sleeping," inviting other users to harass the sleeping avatars (i.e., by stepping on their faces).

In this chapter, we discuss what research reveals about our behavior within VR. We cover what's been done, what the findings reveal about the metaverse and its participants, and where to go from here. We explore two components that make up the metaverse: the people and the environment. First, the people: how does what you look like transform who you are? We consider how changes in *your* and *others*' visual appearance can lead to changes in behavior and beliefs. Second, the environment: how does virtual reality differ from physical reality? What is the technology capable of and how does this contribute to the ways in which we perceive and experience? Lastly, we finish with the limitations of the current literature and what future research is necessary. Taking a step back, this chapter discusses the new moral questions and scenarios that arize almost uniquely within VR, and why it is critical to evaluate the age-old questions on moral behavior within this new context.

## Virtual People: Avatars

**You are what you wear**. Annie is getting ready for a fancy dinner party. She is deciding on what to wear and how to style her appearance. As much as she would like to show up in loungewear, it would most likely not be appropriate for this context. She also knows that how she presents herself is not only for the other attendees at the party, but for herself as well.

She wants to feel confident. She finds herself correcting her hunched back, parting her hair in different ways till she finds that makes her feel the best.

A few weeks later, Annie is getting ready for another dinner party. Except, this time it is hosted in virtual reality. In this preparation process, she has a new set of choices to worry about. She has a wider range of options – clothing options that may not have fit her or may have been out of her budget in the physical world. Here, the concerns that may have concerned her in the physical world – fit, comfort, cost, plausibility, are nonexistent. In this virtual party, she can look like whoever or whatever she wants to. As Annie toggles between the available choices, she realizes that her virtual self looks nothing like she does in the physical world – but does that matter?

Avatars allow for such alterations to be possible. In VR, people are visually represented by avatars – digital characters that most often have anthropomorphic features. Avatars play important roles, as they facilitate the interactions people have with others and the relationships that are fostered in these virtual spaces. Previous research has shown that avatars and how they are designed, including their nonverbal cues (e.g., gestures, postures, expressions), verbal cues (e.g., personalized language, politeness, feedback), and other visual qualities (e.g., attractiveness, anthropomorphism, rendering style, and quality) can affect the impression of trustworthiness, presence, social intimacy, and the like (Donath, 2007; Steptoe et al., 2010). Avatars play a critical role in social interaction, as being able to see what a communication partner looks like results in better performance by reducing ambiguity (Tanis & Postmes, 2007). Moreover, avatars provide different identity cues about individuals, such as their demographics and group membership.

However, in VR, avatars do not always have to match their user's actual appearance. Much like how in online spaces like the Internet an individual does not have to use their real identity (e.g., creating pseudonyms instead of using legal names), in VR one can look like anyone. Beyond modifying surface-level attributes like facial structure, hairstyle, or outfit, a person can change fundamental qualities of the self, such as genders, ethnicities, or even species. However, unlike online spaces where people might be limited to the icons or text-based representations of the self, in VR people can be represented by and embodying 3D avatars. In other words, people *become* the avatar that is representing them. This is a particularly important distinction from other digital spaces like the Internet, as now there is an added element of embodiment. Embodiment, or the psychological connection between a person and their avatar, makes a person feel as though that the body that they are in is their own.

Interestingly, it has been found that people conform their behaviors and attitudes to that of their avatars. According to a phenomenon known as

the Proteus Effect (Yee & Bailenson, 2007), people tend to match their behavior to their digital self-representation, independent of how others perceive them. Across two studies, Yee and Bailenson (2007) set out to see if the attractiveness and height of an embodied avatar influenced behavior. In the first study, participants wore avatars whose faces were either attractive, unattractive, or average. After participants spent a few minutes doing exercizes in front of a virtual mirror to familiarize themselves with their new faces, they were asked to introduce themselves to a virtual confederate. During this interaction, the researchers tracked the interpersonal distance between the participant and the confederate, as well as amount of self-disclosure. Results showed that participants who wore attractive avatars walked significantly closer to the confederate and revealed more pieces of information than those in the unattractive avatar. In the second study, participants wore avatars that were either the same height as, shorter, or taller than the confederate. The participant and confederate would then play a few rounds of a money-sharing task. During these rounds, their negotiation behavior was observed. Results showed that participants who were in the taller avatar were significantly more likely to offer an unfair split than those who were shorter than or the same height as the confederate. Meanwhile, participants who wore shorter avatars were more likely to accept an unfair split than those who were shorter than or the same height as the confederate.

The Proteus Effect has also been tested in other contexts. Other researchers have found that the characteristics of your avatar can affect your dating partner choices, antisocial behavior, food choices, racial bias, financial risk-taking, and consumer choices (for a full review, see Ratan et al., 2019). In other words, *what* you look like matters, because it ultimately shapes what you believe about yourself, how you behave, and how you treat others.

Why do we change our behavior in virtual spaces? There are several theories speculating why we adjust our behavior in such ways. Some believe that it is due to an enhanced effect of self-perception (Bem, 1972), the idea that people develop attitudes by inferring them from observations of their own behavior under circumstances in which they occur. Given that virtual environments are anonymizing in nature, external cues become salient, such as avatars' appearance. People would become particularly sensitive to social cues associated with their identity that they infer from their avatar. Others have suggested that people are primed by external cues, and think and behave in ways associated with their stereotypes (Peña et al., 2009).

What are the long-term consequences of embodying avatars that do not inherently resemble you? As aforementioned, in VR users are not represented by static 2D images or icons. In VR, the user becomes their avatar. Users can see their arms, hands, torso, and depending on the platform,

their legs. And, perhaps if technological advancements go in this direction, we will be able to receive feedback that gives us the sensation of touching our body's skin, hair, and its other textures. How does spending hours embodying multiple identities and adjusting your behavior in intended and unintended ways influence what you think about yourself and the world? In the same vein, given that we can easily switch between features in avatars, do we adapt our behaviors with the snap of our fingers? How do we begin to understand and measure these changes?

This experience of embodying inaccurate avatars may affect different people to varying degrees. Many social VR platforms provide limited customization options, which inevitably leads to inaccurate representations. These limited options can be seen in the lack of options for skin color, facial features, hairstyles and textures, and others. DeVeaux and colleagues (2023) touch on how there is yet to be research that investigates the unique consequences of having limited avatar racial representation in immersive environments. In interviewing users on VRChat, a popular social VR platform that allows users to upload their own avatar creations, they found that there was a dominance in light-skinned avatars and resistance in designing non-white avatars from avatar creators (e.g., due to technical challenges of designing and rendering features such as hair textures). They also reported that avatar choices were shaped by racial harassment, with several Black users hiding their race as a survival tactic. If such trends and preferences shape the avatar economy in social VR platforms, how will expectations and habits that people adopt from such environments manifest in the physical world? We have seen time and time again the issues that arize from being in spaces, communities, and institutions that are dominated by homogenous groups (e.g., people of one race, ethnicity, class, economic class, etc.). However, in VR platforms, the avatar makeup is no longer defined by features that people are inherently born with (e.g., skin color, physical attributes) or factors that cannot easily be controlled (e.g., gender, wealth), but by the designs made by the platform and the trends shaped by the users. In other words, cases of discrimination, biases, stereotypes may be more evident, prevalent, and made consciously.

## Are Avatars People or Objects?

In addition to understanding how embodying avatars affects our behavior, we have to also consider our behavior *towards* other avatars. We know very well that harassment, bullying, and other harmful practices are commonplace in virtual spaces like the Internet. Much like the anonymizing nature of usernames, avatars, and their reduced cues can provide varying levels of anonymity. Avatars can empower us to say things, do things, and treat others in ways that we normally would not in the physical world.

How do avatars promote behaviors that may go against how we normally behave? Are our standards for interacting with avatars different from those we have for physical people? In other words, do we view avatars the same way we view real people? Or avatars lie somewhere in between, closer to objects than people?

Turning to classic theories within communication, Reeves and Nass' Media Equation theory proposes that we respond to media the way we respond to real people (Reeves & Nass, 1996). We respond socially and naturally to media, despite knowing that it is not reasonable to do so. We are able to form social relationships and trust media artifacts, as we perceive them as possessing human characteristics like motivation and personality (Benbasat & Wang, 2005). However, how well does the Media Equation apply to situations in which both communication partners are represented by embodied avatars? It could be argued that embodied avatars are something beyond media and are an entity that bridges the physical world with media. Similar to how we study the properties of the physical world and the unique effects of media with separate lenses, there may be parameters unique to embodied avatars that we are overlooking.

In his book on the mechanisms of moral disengagement, Bandura (2015) lays out the ways in which we convince ourselves that ethical standards do not apply to us in certain situations. Bandura details four principles of moral disengagement: reconstructing conduct, displacing responsibility, misrepresenting injurious consequences, and blaming the victim. The fourth principle asserts that we place blame onto the victim through attribution and dehumanization. This process of dehumanization includes portraying images and descriptions of people who are going to be the victims of one's actions as animals, beasts, or other subhuman objects. This act of dehumanization transforms potential empathy a perpetrator may feel. In social VR, portraying people as avatars is a literal dehumanizing process, as a person and their intricacies are being transformed into pixels with limited cues. A virtual representation of someone cannot be weighted as equally as a real person. In its absolute state, an avatar is neither tangible nor alive – it is a quasi-human being with no soul or physical body (O'Tierney, Kavanagh & Scally, 2019). If the act of representing a person as an avatar is dehumanizing, does this promote the moral engagement Bandura describes, and potentially explain why harassment is so rampant within virtual spaces (Freeman et al., 2022)?

In the same vein, in a study published in 2007, Harris and Fiske investigated how certain social groups are differentially processed in the medial prefrontal cortex (mPFC), which is activated during social cognition tasks. In this study, the authors found that there is reduced mPFC activity in response to people from social groups that elicit emotions such as disgust. This reduced mPFC activity suggests that people in certain groups

are perceived differently, perhaps even less human. The takeaway from this study is that we treat people differently depending how we see and understand them. How do these results translate to avatars and who or what do they represent? It is critical that we understand if we respond to avatars in a way that dehumanizes them. Given that avatars mediate the communication between two parties, if any dehumanization does take place, we need to understand how that influences the interaction.

If it is the case that we fundamentally perceive avatars as something less than human, returning to the Media Equation theory, we are not necessarily treating avatar – the representation of people on media – the same way we would a person in the physical world. In fact, we may be treating avatars as something closer to objects. In a study conducted by Riedl and colleagues (2011), the authors investigated what trust looks like on a behavioral and neurological level when people interact with humans or avatars. In their study, they had people play an economic game in which they evaluated the trustworthiness of different humans and avatars after observing how they handled invested money. Behaviorally, people's trust in avatars was similar to that in humans. However, neurobiologically, the brain distinguishes between avatars and humans. Their results showed higher activation in a brain network that is associated with mentalizing, suggesting that people attribute the concept of a "mind" to humans, but not to the same degree to avatars.

Does this mean that we will mistreat virtual beings? In 2006, Slater and colleagues replicated Milgram's obedience study (1963) in VR. In Milgram's study, a participant is tasked to administer a memorization test to another participant in a separate room. If the second participant – a confederate, unbeknownst to the actual participant – provided an incorrect answer, the participant had to deliver an electric shock, with the voltage of the shock increasing with every wrong answer. With every shock, the confederate would cry out in pain and ask to leave the study, eventually "passing out" as the voltage got higher and higher. When the participant hesitated to continue with the study, the experimenter told the participant that they must continue. The study ended when the participant refused to obey the experimenter or when they reached the highest voltage. The results showed that the majority of the participants obeyed the experimenter and went to the final voltage. In Slater and colleague's VR replication of the Milgram study, the confederate was an agent, a computer-controlled digital representation.[2] In this study, the goal was not to look at obedience (i.e., the experimenter did not tell the participant to continue), but to see if inflicting harm on a virtual being caused anxiety, despite knowing that the virtual being and the shocks were not real. The authors recorded participants' subjective, behavioral, and physiological responses and found that people responded realistically on all three levels toward the virtual being.

There is still a ways to go with researching how we treat avatars differently from how we treat people. While there is a plethora of research on how we interact with various types of media (e.g., see the work done on Nass' Computers Are Social Actors paradigm), there remain many questions on *embodied* avatars. Especially given the critical role that avatars play in social VR settings, we must understand if we view these interactants the same way we view fellow interactants in the physical world. And if not, how our behaviors may change for better or worse.

## Virtual Realism

**Real, enhanced**. Oftentimes people comment on how "realistic" VR is. Here, realism holds many meanings: it can refer to the photographic realism of the scene (i.e., the quality of the rendering), the anthropomorphic realism of the avatars (i.e., the human-like or non-human-like attributes of a person), or the behavioral realism of the avatars (i.e., how much the behaviors of avatars match those expected in the physical world). While there is a consistent sense of realism in the physical world, these varying degrees of realism within VR lead to different ways for us to process virtual experiences.

Consider this scene: Annie is sitting on a bench in a park. Everything around her looks oddly hyperrealistic – she can see every individual blade of grass and how it is shaded, the pores of the person sitting next to her, and the details of the leaves stuck on the windshield of a car passing by across the park. The person next to Annie is looking at her as she talks and has not broken eye contact yet. While the attention is nice, the constant eye contact from this uncanny valley-inducing person makes her shift in discomfort. Annie decides to step away, and as she is doing so, she backs into another person behind her. However, instead of colliding with the person, both Annie and the other person are standing in the same spot. Even weirder, she can see *inside* the person – she can see their teeth, tongue, and eyeballs.

This scenario is not one that is too otherworldly, as in, the people and place were recognizable as ones found in the physical world. At the same time, there were elements of this world and how it operated that were different from what was expected. These differences, some subtle, some large, ultimately lead to affect how one goes about in this alternate world.

Such scenarios are not uncommon in social VR. Many platforms value photorealistic graphics and will put users in environments with incredible details and colors that they may not notice in the physical world but are "forced" to see in the virtual world. Similarly, changes may be made to how other people are rendered, such as how they maintain eye contact or what their facial expressions look like, to enhance user experience.

However, these worlds are not perfect, and bugs may arize: malfunctions in rendering, lack of boundary setups, or badly designed features. Ultimately, these features shape the ways in which people act, think, and navigate in the virtual world.

Such transformations on behavior are unintended, as they were implemented top down by the system and cannot be controlled by the user. These features of enhanced realism, supplemented by bugs and edge cases, lead to changes in how people experience the world. Freeman and colleagues (2022), for example, describe the phenomenon of embodied harassment, a novel form of harassment in virtual social spaces, in which "harassing behaviors are both conducted and experienced through sense of embodiment about one's virtual body, such as a higher awareness of body ownership and more physical and transformative/interactive experiences" (85: 22). The authors describe how in social VR, users are not simply viewing their activities on a screen as in online gaming or traditional 2D-based virtual worlds. Users are actively engaging in the virtual space with both their virtual and physical bodies and bringing in their own history, background, and sociocultural experiences. As Blackwell and colleagues (2019), who are also researchers in the field, put it,

> Users are embodied in avatars that move when the player moves and interact with other players in three-dimensional spaces, enabling violations of personal space and corporeal presence that feel fundamentally different than interactions that occur in other online environments—an experience made potentially more salient by the unique sensation of presence, or the feeling of truly "being there".
>
> (100: 4)

A real case of such enhanced, embodied harassment was described by Zheng and colleagues. In their 2023 study, they describe an instance in which a user fell to the ground after being virtually slapped by another user, suggesting that violence in VR can trigger realistic sensations and elicit real-world responses. Such unintended affordances give rise to novel ways of existing, experiencing, and doing that may be powerful yet dangerous.

In the same vein, there are cases in which users intentionally make use of the designs of a platform to perform behaviors that are either unique to that medium or cannot be demonstrated in the physical world. In the 2023 study described earlier by Zheng and colleagues, the authors analyzed transcripts and comment sections of hundreds of YouTube videos on VR content to observe what kind of social norms form in such spaces and identified safety risks unique to social VR. Of these risks they identify, many include scenarios in which the users make intentional decisions to

use VR to perform morally ill decisions. The authors list examples such as kinetic cues (e.g., poking or trampling a person's face) and proximity cues (e.g., rushing back and forth to crash into people) that users take advantage of to torment others. Likewise, people can use tools within the platform to harass others. In what the authors describe as *virtual crashing*, users use tactics such as adding particle effects like fire spawn animation to cause damage and ruin others' experiences. Given many, if not all, platforms come with some set of tools in which users can spawn new objects into the world, it is no surprize that some users will misuse these tools to terrorize others.

Another example of user-intended harassment occurs through role-plays, the act of taking on and acting out a new personality, values, and goals that are different from your own. Role-playing in social VR is easy, given users can switch between avatar appearances with a few clicks. Zheng and colleagues describe how, in VR, it is challenging to identify the seriousness and intention of the role-players. Other bystanding users may not be able to tell if the role-players are playing around or having genuinely unpleasant experiences. By allowing users to easily put on new identities and completely mask their true self, platforms not only enable the customization of identities, but they also allow users to engage behaviors that would have otherwise not been performed in the physical world.

### Are VR Memories Just as Real?

So far, we've discussed actions that occur within the virtual world. A phenomenon like *virtual crashing* makes sense within the virtual context, but surely such actions do not manifest within the physical world. Why should we be concerned about events that occur inside the virtual world? This raises the question: how do lived virtual experiences change physical behavior, if at all? Beyond eliciting immediate reactions like physically reacting to a virtual punch, can the actions that we perform, witness, or experience influence how we live our lives in the physical world? Can we acquire new knowledge or behavioral changes from virtual experiences that we apply either knowingly or unknowingly into our physical lives? We've seen time and time again from previous research that virtual actions have physical world consequences (for example, prosocial behavior, Herrera & Bailenson, 2021; cultivating empathy and racial awareness, Roswell et al., 2020; climate change awareness, Fauville et al., 2020).

Research hints that VR has the power to alter even our most fundamental moral values. Segovia and colleagues (2009) put participants in VR to evaluate how viewing moral or immoral actions influenced their self-rating of morality and behavior. They showed participants scenarios of an avatar

either distributing first aid kits to avatars or punching other avatars and having the bodies piling up as they fell. Those who saw the immoral behavior were more likely to take more hand sanitizer (i.e., engage in physical cleansing) and rated themselves as less moral compared to those who witnessed the moral behavior. The authors end with the following question: how does viewing immoral behavior in virtual environments affect future behavior in the physical world? Can what we see and believe in VR be powerful enough to reshape our core values on what is good or bad?

One population that may be most vulnerable to this is children. Maloney and colleagues (2020) set out to watch how minors exist and interact in social VR. They observed that children tend to perceive their interactions with other children in social VR as more realistic than those in traditional virtual worlds. They also suggest that children may have difficulty differentiating from the offline and online world, as they are still developing their personality and understanding of the self, others, and the world. Furthermore, other research has shown that it is quite easy to fabricate memories ("false memories") without the consent or mental effort of an individual. Digital media, such as those experienced in social VR, can affect our memory and emotion, and this is especially true for children, who are the most vulnerable (Segovia & Bailenson, 2009). Zheng and colleagues (2023) did find cases in which minors engaged in behaviors such as excessive cursing, which may have resulted from them imitating inappropriate behaviors displayed by adult users. In social VR, children, and adult users co-exist, meaning behaviors that adults engage in without much thought or consideration can be picked up by children and mirrored, and ultimately transferred to the physical world. Despite efforts to protect children in such spaces, it is challenging to stop them from engaging in and witnessing inappropriate behaviors.

In his book *Experience on Demand*, Bailenson (2018) touches on the effectiveness of VR in transferring skills and knowledge. The visceral nature, graphics, and interactivity of VR make it an attractive medium of training. At the same time, this could mean that other skills, such as shooting or punching, may also be transferred. One suggestion is to make VR rules operate differently than the ones in the physical world so that we don't acquire certain skills. Platform designers should consider how we can manipulate how weapons are designed and operated such that, should a VR user encounter a weapon in the physical world, they do not know how to use it. Instead of designing platforms such that users are deprived of certain behaviors, should efforts be made to change how reality functions within VR such that the way it operates is entirely unique and hard to transfer to the physical world? While this is by no means a perfect solution, it is one potential way of approaching this concern.

**What Is the Technology Capable of?**

These harassing phenomena and actions researchers identified are unique within social VR, and their impacts are enhanced by the affordances provided by VR. As technology continues to advance, experiences may become increasingly photorealistic, behaviorally realistic, and immersive. Technological advances in graphical rendering allow for real-time raytracing (Unity), foveated rendering (Patney et al., 2016), and simulations that retain properties in physics (Xiang et al., 2022). Beyond visual properties, there are other sensory feedback that add to a heightened sense of realism and ultimately lead to unique perceptual experiences and behaviors.

In our large-scale, longitudinal study on how avatar appearance and environmental context influence perceptions and behavior, we put hundreds of students inside social VR (Han et al., 2023). For eight weeks, students met in small groups for discussions on a platform called ENGAGE. Hosting hundreds of people inside a shared, networked environment can be challenging, as a lot of unpredictable issues, many of that are unique to VR, can arize. One procedural issue we had to grapple with was audio. Audio can be tricky to do well for several reasons. First, sound gets picked up easily. This means that whatever is going on in the background, from passing cars, phones vibrating, and conversations happening in the hallway, can seep into the virtual world. However, unlike the physical world, in which these sounds are continuous streams with varying levels of volume that depend on distance and location in space, these sounds in the virtual world are somewhat inconsistent, and picked up in pieces. Although a user has the option to turn on spatialized audio, meaning the source of a sound is localized in space, this spatialization is only preserved in the virtual world. Imagine interacting with others in a world and hearing hundreds or thousands of irrelevant sounds blasting through the headset speakers.

During the student group discussions, anyone who was not speaking stayed muted and students who had forgotten to mute themselves and had external audio seeping in were asked to turn off their microphones. Interestingly, unique behaviors arose when students wanted to indicate that they were about to unmute. DeVeaux and colleagues (2022) detailed some of these behaviors that emerged during the VR discussions:

> Other behavioral cues unrelated to real life emerged as well. Muting in ENGAGE requires pressing a button that appears on a person's wrist. Therefore, watching someone motioning their hand towards their wrist was an indicator that they were about to speak and the rest of the class would turn their heads in that student's direction.
>
> (9)

In a non-study context, the reasons audio may be troubling are slightly different. If all audio that is picked up is broadcasted to everyone in a shared space, this means that verbal abuse, slurs, or inappropriate noizes (e.g., moans, grunts) are included. Will others use this as an opportunity to bombard other users with messages or unwanted noizes? If such is the case, how should platforms be designed such that certain sounds are picked up and prioritized? What sounds will make it to the virtual world? Will this influence how people share (or do not) sound?

Second, one of the unique affordances of VR is that it can preserve audio spatiality. When a person says something, you can tell where the sound is coming from and how far away the sound is. This enables multiple conversations to take place without having people having to remain unmuted. This also means that sounds can "sneak up" on a user. Consider this: a person comes up from behind and whispers something close to your ear. This experience may be startling or unnerving. Alternatively, you may have never experienced this. Such behaviors are not commonplace in the physical world. Going up behind people, violating their personal space, and saying something in such close proximity is not a norm. In the physical world, we have an understanding of maintaining distance between our communication partners. In the virtual world, the norms are a little different. Many social VR platforms have designed features that allow users to have a personal bubble that protects them from people coming into their personal space. Multiple people can be standing in the same exact spot, twisted inside one another. Much like how you can experience space violation, you can also experience audio violation. Sounds such as heavy and close breathing can feel visceral and uncomfortable. This feature, much like how space has been used for virtual violence (Zheng et al., 2023), can also be used to attack others.

In addition to audio, there is haptic feedback. Gloves, vests, shoes, arm sleeves, and other wearables have been developed to give users the sensation of touch. By sending electrical pulses across the skin, these wearables allow users to feel sensations ranging from punches and gunshots to caresses and hugs, ultimately creating a sense of realism and immersion. If we can punch people and know that the punches are not "real" but felt, would we do it? If we can push someone off a cliff without them facing physical damages, would we do it? If we could indirectly experience harm without the consequences, what really motivates us to not do so?

As technology continues to advance, more unique ways of knowing and experiencing are being incorporated into VR. We also see advances in other sensory feedback such as smell (see work on olfactory cues by Li & Bailenson, 2017) or taste (see work on thermal taste by Karunanayaka et al., 2018). Such technologies will introduce new ways of experiencing

processes we are already familiar with from the physical world. But how do these virtual transformations transform the way we live within the virtual world?

## The Future of Virtual Cultures

Taking a step back, it is easy and perhaps not fruitful to fall into panic about morality in VR, as there are still many questions that have yet to be answered in this space. What about the children? What about the violence? What about the harassment? The concern is that the metaverse, as it stands now, is under-regulated. As Blackwell and colleagues (2019) assert, because expectations for appropriateness are unclear in these environments, users and moderators are reluctant to make assumptions of people and their intentions. Furthermore, there is no certainty in what kind of user behaviors are to be expected, what kind of consequences those behaviors lead to, and who must be held accountable for those behaviors. This uncertainty stems largely from a lack of understanding of behavior in virtual environments – namely morally questionable behaviors – and what mediates them.

There are a few avenues of research that may help us understand this landscape better. First, further research is necessary on how social VR use changes with time. Curiosity and the desire to experiment with the new ways of existing in VR may drive behavior for novel users. Virtual cultures, however, develop over time. Findings concluded from a single or limited exposure to VR may not meaningfully inform how these virtual cultures will develop. Previous research has shown that time plays a critical role in painting a better picture of how we understand behaviors that take place in VR (for examples, see Bailenson & Yee, 2006; Han et al., 2023). To truly understand the cultures within the metaverse we not only need experiments that collect responses at multiple different time points, but also ethnography that investigates already-existing cultures and how their members, values, and rules have come to be.

Second, we must grapple with the question of *why*. Why do our behaviors change within the metaverse? Are we acting out roles that we have come to learn from the physical world (i.e., the Proteus Effect)? Or are we satisfying a curiosity that would otherwise be impossible in the physical world? Or are we responding to a new set of rules laid out by the environment and the medium? Or are we acting out of courage that anonymity and the ability to change between multiple selves provide within the metaverse?

Third, we need to think about what happens outside of VR. By this, we don't necessarily mean how our behaviors are affected by our virtual

experiences. Rather, what are we doing with this new information we have in our virtual lives? We've now seen that the behaviors enacted, and ultimately the cultures that form, within VR are different. What can we make use of that information? Perhaps one of the biggest differences between VR and the physical world is that the former is entirely created, controlled, and surveilled by programmers, developers, designers, and the other. If surveillance and what we do with people's data is a concern in the physical world, it should be an even bigger concern within VR. Whereas in the physical world, there may be blind spots where a camera isn't present, or a sensor isn't working, or a tracker isn't available yet, in VR, all our movements, physiological responses, and decisions can be easily tracked. Studies have shown that we can identify people with pretty high accuracy by simply looking at how they move (Miller et al., 2020). Beyond movement, there is a lot of biometric data, from eye-gaze, pupillometry, heart rate, and facial muscle movement, that can be collected from these new, advanced headsets that are being continuously released. In the metaverse, where we can become anyone or take on multiple identities (i.e., have multiple avatars), how should a person's profile – or multiple profiles – be created and understood?

## Notes

1 Throughout this chapter the terms social VR and the metaverse interchangeably. Here, both terms refer to a shared social virtual environment accessed through a VR headset in which people interact in real-time and are represented by avatars.
2 An agent is different from an avatar such that an avatar is a digital representation controlled by a person.

## References

Bailenson, J. (2018). *Experience on Demand: What Virtual Reality Is, How it Works, and What it Can Do*. W. W. Norton & Company

Bailenson, & Yee. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(6), 699–716. https://doi.org/10.1162/pres.15.6.699

Bandura, A. (2015). *Moral Disengagement*. Worth Publishers

Bem, D. J. (1972). Self-perception theory. In L. Berkowtiz (Ed.), *Advances in Experimental Social Psychology* (Volume 6, pp. 1–62). Elsevier. http://dx.doi.org/10.1016/s0065-2601(08)60024-6

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3), 4

Blackwell, L., Ellison, N., Elliott-Deflo, N., & Schwartz, R. (2019). Harassment in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–25. https://doi.org/10.1145/3359202

DeVeaux, C., Han, E., & Bailenson, J. N. (2022). Expanding education through virtual reality. In S. P. McKenzie, L. Arulkadacham, J. Chung & Z. Aziz (Eds.), *The Future of Online Education* (pp. 325–336). Nova Science Publishers

DeVeaux, C., Han, E., Landay, J. A., & Bailenson, J. N. (2023, May). A presence of absence: Understanding disparities in avatar racial representation and embodiment in social VR. 73rd Annual International Communication Association Conference.

Donath, J. (2007). Virtually trustworthy. *Science*, 317(5834), 53–54. https://doi.org/10.1126/science.1142770

Fauville, G., Queiroz, A. C. M., & Bailenson, J. N. (2020). Virtual reality as a promising tool to promote climate change awareness. In J. Kim & H. Song (Eds.), *Technology and Health* (pp. 91–108). Elsevier. http://dx.doi.org/10.1016/b978-0-12-816958-2.00005-8

Freeman, G., Zamanifard, S., Maloney, D., & Acena, D. (2022). Disturbing the Peace: Experiencing and mitigating emerging harassment in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1–30. https://doi.org/10.1145/3512932

Han, E., Miller, M. R., DeVeaux, C., Jun, H., Nowak, K. L., Hancock, J. T., Ram, N., & Bailenson, J. N. (2023). People, places, and time: A large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication*, 28(2). https://doi.org/10.1093/jcmc/zmac031

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience*, 2(1), 45–51. https://doi.org/10.1093/scan/nsl037

Herrera, F., & Bailenson, J. N. (2021). Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors. *New Media & Society*, 23(8), 2189–2209. https://doi.org/10.1177/1461444821993121

Karunanayaka, K., Johari, N., Hariri, S., Camelia, H., Bielawski, K. S., & Cheok, A. D. (2018). New thermal taste actuation technology for future multisensory virtual reality and internet. *IEEE Transactions on Visualization and Computer Graphics*, 24(4), 1496–1505. https://doi.org/10.1109/tvcg.2018.2794073

Li, B. J., & Bailenson, J. N. (2017). Exploring the influence of haptic and olfactory cues of a virtual donut on satiation and eating behavior. *Presence: Teleoperators and Virtual Environments*, 26(3), 337–354. https://doi.org/10.1162/pres_a_00300

Maloney, D., Freeman, G., & Robb, A. (2020, November 2). A virtual space for all: Exploring children's experience in social virtual reality. *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 472–483. http://dx.doi.org/10.1145/3410404.3414268

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371–378. https://doi.org/10.1037/h0040525

Miller, M. R., Herrera, F., Jun, H., Landay, J. A., & Bailenson, J. N. (2020). Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports*, 10(1), 17404. https://doi.org/10.1038/s41598-020-74486-y

O'Tierney, A. J., Kavanagh, D., & Scally, K. (2019). Me and my avatar: Acquiring actorial identity. In H. Hwang, J. A. Colyvas & G. S. Drori (Eds.), *Agents, Actors, Actorhood: Institutional Perspectives on the Nature of Agency, Action, and Authority* (pp. 65–86). Emerald Publishing Limited. http://dx.doi.org/10.1108/s0733-558x20190000058006

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics*, 35(6), 1–12. https://doi.org/10.1145/2980179.2980246

Peña, J., Hancock, J. T., & Merola, N. A. (2009). The priming effects of avatars in virtual settings. *Communication Research*, 36(6), 838–856 https://doi.org/10.1177/0093650209346802

Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2019). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect. *Media Psychology*, 23(5), 651–675 https://doi.org/10.1080/15213269.2019.1623698

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People*. CSLI Publications.

Riedl, R., Mohr, P., Kenning, P., Davis, F., & Heekeren, H. (2011). Trusting humans and avatars: Behavioral and neural evidence. 32nd International Conference on Information Systems

Roswell, R. O., Cogburn, C. D., Tocco, J., Martinez, J., Bangeranye, C., Bailenson, J. N., Wright, M., Mieres, J. H., & Smith, L. (2020). Cultivating empathy through virtual reality: Advancing conversations about racism, inequity, and climate in medicine. *Academic Medicine*, 95(12), 1882–1886. https://doi.org/10.1097/acm.0000000000003615

Segovia, K. Y., & Bailenson, J. N. (2009). Virtually true: Children's acquisition of false memories in virtual reality. *Media Psychology*, 12(4), 371–393. https://doi.org/10.1080/15213260903287267

Segovia, K. Y., Bailenson, J. N., & Monin, B. (2009). Morality in tele-immersive environments. Proceedings of the International Conference on Immersive Telecommunications (IMMERSCOM), May 27 – 29, Berkeley, CA, USA. http://dx.doi.org/10.4108/icst.immerscom2009.6574

Slater, M., Antley, A., Davison, A., 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(1), e39. https://doi.org/10.1371/journal.pone.0000039

Steptoe, W., Steed, A., Rovira, A., & Rae, J. (2010, April 10). Lie tracking. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. http://dx.doi.org/10.1145/1753326.1753481

Tanis, M., & Postmes, T. (2007). Two faces of anonymity: Paradoxical effects of cues to identity in CMC. *Computers in Human Behavior*, 23(2), 955–970. https://doi.org/10.1016/j.chb.2005.08.004

Xiang, D., Bagautdinov, T., Stuyck, T., Prada, F., Romero, J., Xu, W., Saito, S., Guo, J., Smith, B., Shiratori, T., Sheikh, Y., Hodgins, J., & Wu, C. (2022).

Dressing avatars. *ACM Transactions on Graphics*, 41(6), 1–15. https://doi.org/10.1145/3550454.3555456

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33(3), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Zheng, Q., Xu, S., Wang, L., Tang, Y., Salvi, R. C., Freeman, G., & Huang, Y. (2023). Understanding safety risks and safety design in social VR environments. *Computer Supported Cooperative Work*, 7(CSCW1), 1–37. https://doi.org/10.1145/3579630