Revista Română de Filosofie Analitică,
Romanian Journal of Analytic Philosophy
Vol. XVII, Iulie-Decembrie 2023, Nr. 2, p. 7-31
ISSN (ediția electronică): 1843-9969,
ISSN (ediția tipărită): 1844-2218
DOI: 10.62229/rrfaxvii-2/1

EMBODIMENT AND EMPATHY TRAINING

ALEXANDRU PETRIŞOR[1]

Abstract: The "Empathy Machine" hypothesis posits that virtual reality (VR) environments can significantly increase empathy more effectively than any other media. This has been widely debated, empirically tested, and proposed as a non-invasive alternative to traditional forms of moral enhancement. In this article I argue that: 1) VR has the potential to enhance empathy at least as effectively as other media, such as cinematography and literature; and 2) currently, VR does not achieve this potential. This challenges the notion that VR is the most effective medium for empathy enhancement. Following a review the literature supporting and contesting the efficacy of VR for empathy enhancement, I provide empirical and theoretical arguments to substantiate the claim that current VR projects fail to enhance empathy, and that this is due to misconceptions about embodiment in fictional settings and about different types of immersion. I conclude by suggesting that VR experiences could be as effective as other media in enhancing empathy, but present VR applications overemphasize participatory immersion at the expense of narrative and representative immersion, thereby limiting their effectiveness. The discussion underscores the need for a balanced approach to VR design that incorporates the successful elements of other media in fostering empathy.

Keywords: empathy, virtual reality, embodiment, make-believe, moral understanding.

[1] Alexandru Petrişor is a doctoral student within the Faculty of Philosophy at the University of Bucharest.

Introduction

Let us call the "Empathy Machine" hypothesis the claim that by using virtual reality (VR) environments we can significantly increase people's empathy much more effectively than with any other media. This hypothesis has received much discussion in the literature, being the test target for many studies and being seen as a non-invasive alternative to other forms of moral enhancement (Zahiu et al., 2023). In this article I will argue: 1) that VR could enhance empathy at least as well as other media (cinematography, literature, etc.), and 2) currently this is not the case. (1) establishes a lower limit for well-constructed VR experiences that it shares with other media, while leaving the door open for it to be even better. (2) serves to deflate the premature optimism of the 'Empathy Machine' hypothesis. In the first part I will elucidate each conjunct of my thesis. The first one is the target of many articles in the literature (Zahiu et al, 2023; Bloom, 2017b; Lara & Rueda, 2021; Rueda & Lara, 2020; Sora-Domenjó, 2022; Martingano et al., 2021), although the formulation can differ depending on what one takes to be the case. If I think that VR is an empathy machine, then I will argue for a stronger variant of (1): that VR is better than other media at enhancing empathy. Alternatively, if I believe that VR is an inadequate medium for enhancing empathy, then I will argue that it isn't nearly as good for said purpose as other forms of media. There may be other possibilities but the discussion will reveal that (1) is the best formulation of this thesis because it keeps VR in line with other media. Regarding (2), my argument will be based both on empirical findings and on armchair theorizing about how current VR projects miss the mark when trying to enhance empathy. My speculative suggestion is that this failure stems from misconceptions of both embodiment in fictional settings and immersion. This will work as a how-possibly explanation of why VR isn't the empathy machine many authors expect it to be. As such, I will try to better articulate what are the relevant factors that go into empathy enhancement via VR experiences.

The article is divided into five sections. The first one defines the terms that I have hitherto left to their intuitive meaning in order to better understand the issue at hand. Also in this section, I will comment more on the motivation behind philosophical discussions about empathy enhancement. The next section deals with some familiar accounts from the literature on VR and empathy. The third section contains the arguments supporting my thesis, dealing with (2) first and (1) subsequently. As we will see, the reversal is well-motivated. In the fourth section, I will reject several counterarguments to my thesis. The fifth section concludes the article.

1. Defining VR and empathy

1.1. Virtual Reality (VR)

Since Virtual Reality is so popular both in recent academic literature and because of devices that make their way into the public's hands, I need to clarify what I mean by VR. We can describe VR as a medium that uses computer modeling to simulate 3D environments with which the player/agent can interact. I will focus on HMD VR devices (head-mounted displays), but the discussion isn't limited to this type of technology. This step just lets us more easily refer to current widely available technology. A plethora of VR studies use such an interface (Martingano et al., 2021), but an HMD does not differ fundamentally (although the effect on the feeling of being there in the virtual world is magnified by its use) from a simple computer screen – and this may matter when taking into account how interactive movies function. Nonetheless, my arguments are not limited to HMDs.

One might wonder what makes a VR experience fundamentally different from a video game played on a regular display. Here, Chalmers' observations can help: "What is distinctive about VR is that its virtual worlds are immersive. Instead of showing you a two-dimensional screen,

VR immerses you in a three-dimensional world you can see and hear as if you existed in it" (Chalmers, 2022, p. xii). Echoing our definition, Chalmers describes VR environments (there is some difficulty in extending this definition to VR experiences, as I will later suggest) as immersive, interactive, and computer-generated. So, immersion is a key component of VR experiences, inasmuch as they presuppose VR environments. From Chalmers' quote, we can gather that the relevant sense of immersion is perceptual: the fact that you can see mountains and hear the gust of wind is what makes it an immersive environment. As we will see, this is just one type of immersion to consider when talking about experiences in VR environments.

Presumably, by using computer modeling you could, in an intuitive sense, be put in the shoes of any other being with similar enough perceptual systems. For example, by the use of VR, you could be in the shoes of a Syrian refugee (Arora & Milk, 2015), or a black man in some unspecified US city (Cogburn et al., 2018). It doesn't seem to be the case that you could be put in the shoes of a bat if that would mean getting to know what it is like to be a bat (Nagel, 1974). Being in someone's shoes is not that clear; in the third section, I'll advance several ways to cash it out.

1.2. Empathy

Many definitions of empathy are available. In different corners of the literature, empathy gets assigned different roles/importance: as a way of getting to know what somebody feels, as emotional contagion, as getting to know *what it is like* for someone to feel something, etc. (Gallagher, 2020; Zahavi, 2014). It would help to use a definition that doesn't presuppose too many specific choices about the role and importance of empathy. This is why I follow Zaki and Ochsner, taking empathy to be "the ability and tendency to share and understand others' internal states" (Zaki & Ochsner, 2016, p. 871). So construed, empathy can be described as having three components:

1) experience sharing = the tendency of perceivers to take on the sensorimotor, visceral, and affective states of the targets;

2) mentalizing = perceivers' explicit reasoning about targets' internal states using lay "theories" about how situations produce internal states;

3) prosocial motivation = through which individuals who share and understand targets' states often are compelled to help those targets. (Zaki & Ochsner, 2016, pp. 871-872)

The definition of empathy describes empathy as both ability and tendency: (1) is a tendency, whereas component (2) is an ability. This implies empathy is not a "success term" (Walton, 2015, p. 2). You may empathize with X but wrongly attribute to her the mental state Y, instead of Z. Presumably, your failure to properly empathize would stem from your lack of ability in empathizing. Maybe you need more training: the empathy trainer may prescribe reading more novels or watching more movies. (This raises questions regarding the individuation of mental states and attitude ascriptions, respectively.) Perhaps terms for abilities can sometimes be success terms too – if you fail to initiate the process at all; no representation means neither success nor failure. As you grow more competent in your empathizing skill, you gain the ability to attribute the correct mental state in cases where previously you would have been silent. Even so, the point of characterizing empathy in dispositional terms is that it can be improved. Given empirical data supporting this hypothesis (Persson & Savulescu, 2018), I bracket worries about the most intuitive way of characterizing empathy and accept the given definition as a working one.

Although (2) explicitly mentions lay "theories", we should not take this to mean a commitment to theory of mind (ToM) accounts. I use "mentalizing" to refer to whatever mental processes that help us attribute mental states to others, in ways *somewhat* linked with experience sharing and motivational aspects of empathy. Using "mentalizing" this way sets aside cases of experts at attributing mental states to others who lack the

motivation to help them and/or appropriate affective reactions to moral violations (such as anger/disgust towards harm done to innocents). Such cases are paramount in evaluating an individual's specific moral profile, but fade when we consider moral behavior on average, in the general public.

2. The Empathy Machine

Having fixed working notions of empathy and mentalizing, I now turn to the imaginative aspect of empathy, the focus of debates surrounding empathy enhancement. Intuitively, some people simply fail at imagining how others feel and therefore fail to understand their mental states. This, in turn, can lead to disastrous consequences. Zahiu et al. put the problem this way when they distinguish between bounded and reflective empathy (2023, p. 4), which correspond to experience sharing and mentalizing, respectively. Bounded empathy is an evolutionary strategy of behavior optimization that is adaptive to the challenges of a given environment. It is triggered automatically and spontaneously biased (see Bloom, for more on the biased aspects of empathy). Reflective empathy is described as being "supervised by voluntary acts of imagination and deliberation" (Zahiu et al., 2023, p. 5) and thus under cognitive control. Is there room for improvement on the experience sharing/bounded empathy side? Maybe, but it is hard to see how, given its evolutionary design and biases (Bloom, 2017a). Since humans have a lot of cognitive resources that can be leveraged, and since training can help improve many other cognitive-based abilities (such as playing chess), we might have better chances at improving our mentalizing skills, rather than our experience sharing ones:

> "What VR technology can do is to scaffold our imaginative powers. VR technologies are uniquely equipped to attach vivid sensory representations to experiences and interactions. It can bring certain experiences in the here and now, instead of letting individuals rely on imagined hypotheticals. This, in turn, has the potential to enlarge our moral understanding." (Zahiu et al., 2023, p. 6).

Let us now use this passage as representative of empathy enhancement with VR, and refer to it in a short-handed manner as "(EM)". It is not clear if (EM) targets experience sharing or mentalizing. On the one hand, it seems to be explicitly about mentalizing, since it involves imagination as used for understanding the mental states of others. Presumably, this is to be done by imagining oneself in the shoes of the target. On the other hand, (EM) claims that, by making people subject to vivid sensory representations of certain experiences/interactions, we avoid the need to imagine said circumstances. On the supposition that mentalizing requires imagining, or at least some degree of deliberation, it seems that VR eschews both. If so, the person connected to the empathy machine does not mentalize anything, but instead shares an experience with some target. This does not mean that (EM) is defective, since the relevant VR experiences might not involve representations so vivid as to forestall the need for any mentalizing or deliberation. But this is an empirical and technological question, one that can be probed by looking at studies and checking whether VR training increases empathy (specifically via which component), and how data might vary with newer generation VR.

This is very important for my thesis. Since empathy is described as having three components, it is at least conceivable that each can operate in the absence of another. This isn't entirely true if one looks at the neural correlates of said components, even though the neural circuitry differs significantly with respect to the first two (Zaki & Ochsner, 2016). One might have a pronounced tendency to share the experiences of others, so she will naturally pick up the mental states (emotions/moods) of other people when they are in her proximity. It is at least conceivable that such a person could perform poorly on mentalizing tasks, especially ones that require her to empathize with people not in her proximity and differing in some non-essential characteristics enough so as to not be perceived as a member of an in-group. Evidence from studies also suggests that this is not just a possibility, but is the actual case for some people (Zahiu et al.,

2023). Likewise, one might be a master mindreader even though she is not able to easily engage in experience sharing. Evidence suggests this is the case for some (Rijnders et al., 2021). Since these two can come apart to some extent, we can raise the question of what exactly it is that VR experiences train.

Compare: literature, too, tends to increase mentalizing skills, as opposed to increasing the disposition to share an experience (Martingano et al, 2021). What (if anything) makes VR (but not old-fashioned reading) an Empathy Machine? Bounded empathy is prone to biases, corrigible if people rely more on reflection. But if VR does not supply us with the right increase in the right parameter (e.g., reflective empathy) then its label as an Empathy Machine seems unwarranted. Worries also linger about the ethics of using VR experiences for increasing empathy:

> "The pleasures of toxic embodiment offered by witnessing racial suffering in VR extend these precarious conditions of life. Rather than trying to automate compassion for those who suffer as a result of racialized violence, indifference, and hyperautomation, we need forms of critique that show us how emotions like empathy and compassion have *alibi*-ed untenable material conditions of labor for racialized and gendered people long before VR claimed them." (Nakamura, 2020, p. 61).

Nakamura targets VR experiences that aim to put a player in the shoes of a minority. Not all VR experiences aim to do this and there is no requirement that they should, even though it seems to be an easy choice to make given the way the world is right now.

These discussions lay a strong emphasis on embodiment and on identifying the virtual agent with the person undergoing the VR experience. You are supposed to be in the shoes of a refugee and that means that you are supposed to (fictionally) have such-and-such characteristics. This is facilitated by vivid representations you are supplied with. But, we might ask, what game of make-believe is played when engaging with such fictional worlds? The lack of this kind of analysis is felt on both sides. Presumably, Nakamura would not have any

problem with literary works of marginalized people describing their first-person experience. Is this to be explained by virtue of the literary medium itself, or by features of the types of games of make-believe that one engages in when reading said works? On the other hand, is empathy enhanced by vivid representations, or in virtue of the game of make-believe that VR experiences authorize? These questions shift focus away from the surface characteristics of the VR medium to the way in which users participates in it.

An analogy might help: the way the current discussion goes is akin to saying that paintings done in a realistic style are better at making one understand the world of a painting than paintings done in an impressionistic style. This ignores the onlooker's role in engaging in games of make-believe with said paintings, and as a result one might misidentify the parameter X (let us say proficiency in understanding paintings for the sake of the example) responsible for the increase in skill Y (understanding the fictional world of a painting) with some characteristic related to Z (painting style), that nonetheless is not essential to X or Y.

(EM) is also a thesis about a type of understanding: moral understanding. Empathizing with someone not only helps get a sense of how that person feels but also 1. functions as a motivator for doing something about it (to remove the cause of harm) and 2. gives rise to first-hand propositional knowledge (or know-how if you are an anti-intellectualist when it comes to skills and abilities) that is then used in moral and action-guiding inferences, knowledge that is then applicable in multiple structurally similar cases. You gain moral understanding when you have "a grasp of the connections between moral reasons and moral conclusions" (Hills, 2020). As construed here and as I believe that (EM) is intended, VR helps develop moral understanding because it targets the cognitive aspect of empathy. This is what makes possible the multiple applicability of empathy-triggered inferences (that the same

harm is being done in cases of torture even though the subjects have different skin colors).

Remember that we assume, in line with Zahiu et al., that bounded empathy is biased and unlikely to be corrected. What makes one grasp moral knowledge should enable assenting to the moral structural similarity of cases X and Y, differing only in characteristics such as the skin color of the victim. It also should also allow one to generalize their findings: after being exposed to the harsh life of a refugee, one should draw universally quantified conclusions, that all such acts are bad, etc. One could then apply this new-gained knowledge and understanding so as to deal with new cases. In contrast, bounded empathy limits itself to one individual or one group. This is why we should focus on the measurable increase of cognitive empathy (mentalizing). As we will see, there are strong reasons to doubt that current VR experiences are able to fulfill (EM). And this is to be explained, I contend, by employing Walton's (1990) influential analysis of fiction and representational works of art.

Empirical data analysis does not support (EM) unqualifiedly. One recent meta-analysis has found that VR interventions improved bounded empathy/experience sharing, not reflective empathy/mentalizing (Martingano et al., 2021). This meta-analysis surveyed 43 studies with a total of 5,644 participants, so we can take it at least as a worrying result for (EM). One point worth mentioning is that the authors also included presumably less immersive interactive movies. This could be seen as a limitation. The authors, however, found that "more immersive delivery devices that used head mounter displays did not have a significantly larger effect on empathy than non-immersive delivery devices that ran on a normal computer desktop or headphones" (Martingano et al., 2021, p.13). It seems unlikely to say that VR could be an empathy machine and yet fail to increase the ability to mentalize. It seems even less likely that VR could be such a machine in virtue of its immersive embodiment and vivid representations and yet fail to elicit better results than interactive movies which have a lower degree of bodily immersion.

I aim to argue that VR experiences could be just as good as other media for increasing empathy (focusing on the mentalizing aspect) but currently, fail to realize that potential. This meta-analysis supports the second conjunct. I now turn to elucidating the important characteristics of the games of make-believe that VR experiences authorize in order to support the first conjunct.

3. Embodiment and immersion

VR experiences differ from those in other relevant media given immersion and the feeling of embodiment VR procures (Chalmers, 2022). Immersion and embodiment *cannot* be used interchangeably, a lesson sometimes ignored (e.g., Sora-Domenjo, 2022, p. 6, which contrasts immersive experiences with 2D movies). I hold that most current VR experiences lack a focus on narrative immersion[2], as opposed to embodiment (perceptual/spatial immersion) and a lack of agency given to the participant in such experiences. In Walton's (1990) terms, a player/participant who engages in a VR-authorized game of make-believe may nonetheless lack representational capacities needed to explain their actions and *de se* imaginings. In the following subsections, I will argue that embodiment does not guarantee immersion and I will illustrate by means of VR experience samples the ways in which they fail to prescribe the adequate *de se* imaginings that will engage mentalizing.

3.1. Four dimensions of immersion

Being immersed in something that can be described as a state in which one 1) is attentive to the activity in which one is immersed, 2) fails to pay

[2] As will be detailed below, a VR experience is narratively immersive if and only if the subject's experience in VR has a coherent narrative structure and flow. To illustrate with a counterexample: if your favorite historical movie abruptly interposed the story of Spider-Man halfway through, it would lack narrative immersion.

attention to her surroundings if these don't pertain to the activity she is immersed in. I take immersion as a property of one's engagement in an activity. This should not be controversial, as one could define derivative ways in which they are immersed in, for example, a culture by reference to activities pertaining to said culture. Being immersed in a fictional world would then mean that the player pays attention to the game of make-believe she takes part in.

I treat two VR experiences as a sample representative for the whole set of currently available VR experiences which could be used to enhance empathy. One is *Clouds over Sidra*, an interactive movie in which one is presented with the life of Syrian refugees, and the other one is *1000 Cut Journey*, in which one is put in the shoes of a black man as he experiences instances of racism. I choose these two thanks to their popularity in the academic literature and beyond. *Clouds of Sidra* is an interactive movie, but that should not detract from the points I will make, since it functions almost exactly like *1000 Cut Journey*, with the exception that you cannot throw a basketball at any point in the experience. Should that make for a qualitative difference between the two? I see little reason for this claim. *Clouds of Sidra* asks the participant to follow the speaker, and this is done by turning your head around if you are using an HMD. So, there is still a feeling of embodiment/presence in the experience, even though it isn't as complex as in *1000 Cut Journey*. Both experiences allow the user to participate by moving their head in inspecting the virtual environment. In both cases, the characters with which they interact are recordings of real actors playing their designated roles. As such, there isn't much interaction between the player and the non-playable characters, but there is a high representational fidelity of said recordings as props. Due to the likeness that they bear to real experiences, they can easily prompt imaginings regarding their facial expressions or body language. Presumably, this was meant to avoid the uncanny valley effect, which might have influenced empathic responses. Nonetheless, representational accuracy (of people, of interactions) trades off against the level of participation the player exhibits.

The trade-off matters. If the VR experience supplies me with the highest-fidelity sensory information I might feel immersed in the sense that I find it difficult to come to believe that what I am experiencing is not the real world but just computer-supplied images. But if I cannot engage properly with said world – if I cannot walk over to the basketball course and throw a basketball, if I cannot talk to other characters within the world – then I will find it less difficult to come to believe that I am only in a VR experience[3]. Immersion should, then, not be left to intuitions unchecked by conceptual analysis. Balcerak Jackson & Balcerak Jackson aptly note: "There is nothing wrong with such metaphorical characterizations, of course, but their utility is limited when it comes to thinking systematically about VR and our engagement with it" (2023, p. 19).

To illustrate, Zahiu et al. seem at first blush to equate immersion with the feeling of embodiment (2023, p. 3). Likewise, Zaki contrasts VR with other "non-immersive" forms of media, like literature (2014). Nakamura (2020) views VR as allowing voyeuristic engagement; I take this to mean that it does not fully immerse oneself in the experience of a marginalized individual. Bloom (2017b) also emphasizes this lack of full immersion and its replacement with a voyeuristic type of behavior. Such stances err in taking immersion to consist solely in the feeling of embodiment that a participant has when engaging in VR experiences.

(Balcerak Jackson & Balcerak Jackson, 2024) distinguish four different types of immersion: 1) representative immersion; 2) participatory immersion; 3) affective immersion; and 4) narrative immersion. A VR experience is representationally immersive iff it involves a rich and coherent network of perceptual or affordance representational states. VR experiences with photorealistic graphics should rank high on representational immersion. The affordance condition covers more complex mental states that should be prompted by the experience. For instance, being in a VR environment that resembles a city should generate specific representations/affordances (e.g. crossing

[3] I will leave aside differences between expert VR users and novice ones (Chalmers, 2020).

the street to get to the bakery). Balancing on a plank (in a VR environment) generates the state of believing that one is going to fall if one were to step on the side of the plank.

A VR experience is participatorily immersive iff the participant can act so as to determine what the virtual environment is like and how events unfold in it. So defined, participatory immersion differs from what Chalmers calls interactivity. *Clouds over Sidra* is still participatorily immersive because one can choose how to position themselves in the scene and where to focus their attention. This can be extended to literary works: you can choose how to imagine some details from how a room is described (as they are not determinate enough to mandate one single interpretation), therefore you participate in the creation of the fictional world that you are engaging with.

A VR experience is affect immersive iff one emotionally or affectively participates in the virtual experience. This is a stronger condition than merely requiring that a user be psychologically participating in the virtual experience. Imagine disinterestedly playing a VR game in which you have to jump platform to collect coins. Little if any emotion is felt whilst doing the boring task. Nonetheless, one still believes that virtually, there is a platform with ten coins over there, and that after collecting said coins one can proceed to the next level.

Lastly, a VR experience is narratively immersive iff the subject's experience in VR has a coherent narrative structure and flow. This is hard to make precise given the many different types of narrative structures out there. The important point is that this is a type of immersion that differs from the previous three. As we will see, narrative immersion is what many VR experiences ignore.

3.2. The narrow focus on embodiment

My argument for why many current VR experiences fail to deliver improvements in mentalizing is the following:

1) Current VR experiences are developed with the assumption that embodiment guarantees immersion in the relevant sense.

2) Current VR experiences are developed with the assumption that immersion in the relevant sense guarantees improvement in mentalizing.

3) Embodiment does not guarantee immersion in the relevant sense.

4) Therefore, current VR experiences do not improve mentalizing (thus, explaining the empirical data).

Obviously, this is a schematic version, as there isn't such a thing as immersion without a work in which a participant is immersed, and that work has to have certain characteristics in order to target mentalizing. Appropriate kinds of immersion also need to be specified. Nonetheless, this formulation is enough to get the point across. Pending filling in the details, the argument seems valid. I will primarily consider the set of immersive experiences that improve mentalizing (so I eliminate player experiences of immersion in the role of a super-soldier shooting aliens). "Immersion in the relevant sense" sets aside works that only give a representative sense of immersion, which is unlikely to improve mentalizing. Similarly, works that only target affective immersion might improve bounded empathy but not reflective empathy. Nobody would hold that a VR experience with photo-realistic graphics, but nothing to do with it but walk around, is relevant to developing empathy. (1) and (2) come from my reading of the literature and I already presented evidence for their plausibility. I will therefore argue for the truth of (3) in order to show the argument is sound.

Friends and foes of VR as a medium for empathy enhancement presuppose that immersion has to do with the sense of embodiment one experiences when putting on an HMD. This is only partially correct. It is more immersive to have a first-person perspective in a VR environment than engaging with it via a detached computer screen, all other things being equal. However, the *ceteris paribus* clause obscures the complex interaction between different aspects of a VR experience that help foster the sense of immersion. In our sample of VR experiences, even when there is a strong sense of embodiment, either the experience elicits emotional

responses for other reasons (the sad music present in *Clouds over Sidra* and the performance of actors in both this experience and *1000 Cuts Journey*) or the experience fails to elicit emotional responses because it neglects other relevant factors; for example, how *1000 Cuts Journey* undermines its own goals by failure to build narrative coherency and not letting the participant engage in certain types of mental actions that would foster mentalizing and a sense of representative immersion.

A VR experience isn't identical to a VR environment. The former designates the authorized games of make-believe to be played with the VR environment, whereas the other only points to a set of props that are to be integrated into that VR experience. Moreover, the VR environment can include elements that do not feature in the VR experience. Does representative immersion enhance empathy? Surprisingly, this is often taken for granted. Your empathizing with the refugees *in Clouds over Sidra* owes a lot to the sad music playing in the background and the authentic depictions of actual refugees. Being in the shoes of a refugee doesn't mean that you hear sad violin music throughout. Sad music helps form affective immersion but it sacrifices representative immersion.

It is not fictional that the refugees (including the player) can hear the sad music, therefore the music isn't part of the VR experience, even though it is part of the VR environment. Hearing sad music when witnessing people gathered around a table is more likely to elicit imaginings about funerals than hearing upbeat happy music when presented with the exact same scene. This one aspect perfectly illustrates that the VR environment helps frame the VR experience by setting up an intended interpretation (the authorized game) of what is to be fictionally felt and believed.

But if our goal is to enhance cognitive empathy, having the right music playing in the background is similar to solving a test and having the answers written on the back of the page. It leaves little room for the participant to figure out for herself what those fictional people are feeling. In this sense, it comes across as second-hand knowledge: sad music works as a testimonial basis for forming the belief that what the characters are

experiencing is sadness. It gives you a reason to believe something, as opposed to merely prompting you to inquire into what is fictionally true.

Do real people in VR enhance empathy and moral understanding? Or, rather, recordings of real people, as opposed to digital characters that are coded to respond to what the participant does. Their facial expressions help trigger emotions in the participant, but once again little room is left for mentalizing when you can directly see sadness on display. Add to this the disembodied narrator's voice present throughout, and one cannot help but be reminded that one is experiencing a performance put on display for her/him.

In VR experiences such as our two samples, disembodied narrator voices function as an egregious case of an "aside to the audience" (Walton, 1990). This breaks immersion (all kinds) by making the participant aware that one is playing a game of make-believe with a certain work. In the words of Chasid, one's experience with the fictional world becomes "doxastically mediated" (2023). The participant becomes aware of the rules by which she is playing the game of make-believe which is her VR experience. To illustrate, at the beginning of *Clouds over Sidra*, one is directly addressed by Sidra, the girl refugee who is the main character in the work. Then the perspective shifts to different scenes showing a day in Sidra's life, and the participant can only hear her disembodied voice whilst witnessing said scenes. If, in the beginning, one could have a sense of deep immersion (narrative but not only) generated by implicitly understanding what is presented as fictionally true and, therefore, to be imagined, later the rules of what is to be imagined become opaque and doxastic mediation sets in. Is it fictionally true that Sidra walks you through these various scenes, or are you presented with what you are supposed to imagine whilst talking to her?

Other factors may help you stay on track (the narrative component of immersion is developed through Sidra's narration) and affectively immersed in the experience, but the fact that you embody a refugee certainly is of no help here. A character could be facing you and talking to you, the audience member, even if you are watching it from a movie

screen, not an HMD. Psychological participation is still high even absent a sense of embodiment. Embodiment does not guarantee immersion: there is an acute need for other factors as well (coherency of narrative, promptings for affective responses and inquiries into morally relevant factors).

Thus far, I focused on *Clouds over Sidra*. But *1000 Cuts Journey* makes more serious mistakes. Once again, you are presented with a disembodied narrator voice that generates doxastic mediation of what is to be imagined. The feeling of embodiment is more acute: at one point you are in front of a mirror, at another you can throw a ball and some cubes. All this is insufficient for the relevant sense of immersion that enhances mentalizing, for embodiment could have equally been experienced this way in a game about jumping on platforms to collect coins. In terms of narrative coherency, the experience is lacking.

In the end, you are given a moral lesson from the narrator's voice saying that this is the experience for many people in the world and that they hope you leave the VR experience being more open to evidence of racism. Literary works do tend to display sentences such as "X is bad", but there is a way of doing it that does not come across as moralizing or as giving away the answer to the test question (as second-hand knowledge). There are many ways to avoid this, but they essentially involve *propagating*, not *transmitting* knowledge. Knowledge is transmitted (by testimony) when "the speaker's knowledge that *p*, expressed through her testimony, is an epistemic ground of the learner's knowledge that *p*" (Hills, 2020, p. 401). In contrast, propagation of knowledge takes place when "the speaker's knowledge that *p*, expressed through her testimony, is a causal factor and an epistemic influence on the learner's gaining knowledge that *p*, but not an epistemic ground of it" (Hills, 2020, *ibidem*). It is important to distinguish transmission from propagation, as the first one yields second-hand knowledge, whereas the latter, by prompting the learner to herself go through the process of gaining the knowledge that *p*, yields first-hand knowledge.

The participant in a game of make-believe is a reflexive prop and generates fictional truths about himself (*de se* imaginings). It is fictionally true in the game of make-believe that I play with *1000 Cuts Journey* that I am a black man and live through such and such experiences. Moreover, I imagine this *from the inside*. I do not merely imagine that I see a police officer, I am imagining *seeing* the police officer. Embodiment is an important aspect of immersion because it prompts imaginings *from the inside*. You get to imagine *throwing* a basketball, not just that you throw it. But that is not enough. Disembodied narrators transmitting knowledge rob the participant from thinking through/inferring/inquiring about the situation they are in, despite feeling at all times embodied. Lack of representative immersion undermines enhancement in mentalizing.

The way moral knowledge is transmitted in *1000 Cuts Journey* also emphasizes the experimental setting in which these experiences are had: if immersion is what you are after, why would you emphasize that what the participant has just witnessed is a VR experience done at X university under the supervision of Y professors and researchers? I take this to suggest a lack of focus on the narrative the participant is supposed to experience. The participant in the *1000 Cuts Journey* authorized game of make-believe is faced in 10 minutes with three different episodes from the life of supposedly the same man. No other character makes more than one single appearance. What ties together these three episodes is the narrator giving prescriptions to imagine these as episodes of the same man. There is barely enough time to think about what is happening before you are given the imperative to combat racism. This is clumsy storytelling, and it shows in the fact that participants are not left with improvements in mentalizing skills after undergoing the experience (as per section 2). The feeling of embodiment is present throughout, but because of a lack of narrative immersion and representative immersion when it comes to the sorts of mental actions that a participant can fictionally go through (inquiring whether how skin color could ever warrant bad behavior from others), empathy enhancement and moral understanding are undermined.

We now have good reason to conclude (3) is true. Mere embodiment does not guarantee the intended sense of immersion that would foster improvements in mentalizing. There is little reason to believe that VR experiences are much better at this task than any other media – unless one wants to argue that people are more prone to undergoing VR experiences than anything else, which is an empirical question, and intuitively the answer is negative. The truth of (3) fits well with the data presented in section 2 and it also has the merit of identifying what relevant factors go into making a VR experience that could improve mentalizing. Only those factors are precisely the ones that VR as a medium shares with other media: narrative coherency and the ability to prompt *de se* imaginings that would foster moral understanding.

4. Replies and objections

I have assumed that developments in mentalizing depend on a certain level of narrative and representative immersion. Is the assumption unwarranted? Mentalizing has been much discussed in child development studies and social cognition, and might even be a misnomer (some ToM critics would charge that we rarely mentalize [Gallagher, 2020]). I stay neutral on how to best construe ToM, and on which account of social cognition is preferable. The foregoing does not presuppose a ToM-based approach: 1) direct social perception is not ruled out, being available for some sorts of VR experiences (such as those with live-action characters); 2) my view explicitly deals with narratives and particulars, not just with generalizations and reasoning about instances based only on inferential rules; 3) any emphasis on inferences, in order to cash out a moral understanding thesis, goes well beyond ToM and social interaction theories, since moral problems go beyond simply knowing how to interact with others. As such, my account is pluralistic about social cognition. Whether you end by call narrative-based inferences distinct in

kind or not from theoretic inferences does not undermine my discussion of VR empathy enhancement. I explicitly focused on representative immersion regarding the sorts of mental acts that people undergo when coming to believe that "X is bad" because moral understanding presupposes such acts and is essential to developing the mentalizing aspect of empathy. Narrative immersion was emphasized because we do not attribute mental states to people in a void, but always in a social context, relating them to their goals and moral evaluations besides many other factors. Add to this that engaging in games of make-believe and other such pretend play is a great opportunity for self-knowledge (via *de se* imaginings), and the necessary links between narrative coherency and developments in mentalizing and moral understanding via VR experiences (or any other media for that matter) become visible.

How is a participant supposed to develop, on my account, the necessary know-how to reach a moral understanding? Presumably, one has become an expert empathizer with respect to mentalizing. VR experiences might afford an efficient know-how knowledge transfer, thanks to their specific medium characteristics. I take no stance on this issue, either intellectualist or anti-intellectualist, as I believe that both could be easily integrated into discussing the different types of immersion and how they contribute to empathy enhancement. One good argument in favor of my account, as opposed to one that emphasizes the essential know-how that comes with the agentive aspect of VR embodiment, is that mentalizing is significantly improved when one reads novels or watches movies or plays. There is no embodiment there, but still one either acquires the propositional knowledge that such-and-such people feel such-and-such emotions when they find themselves in certain situations or, alternatively put, one still gets to know how to react in certain situations based on what one has fictionally experienced.

One could say that I have not proved that mere embodiment does not guarantee immersion in the relevant sense, but instead, I only managed to prove that these samples do not push embodiment to its full

potential. Embodiment is linked with a sense of agency. Being an agent presupposes certain actions being available, in the context of the narrative, to the participant. These samples do not provide said affordances, therefore they should not be brought up when discussing the merits of VR as an empathy enhancement machine. This transcendental argument of sorts against my position fails because it leaves the notion of agency completely open to interpretation. When are we expected to feel like a proper agent in a VR environment? How big is the set of actions that should be available to the player? Should a player be able to role-play anything in a VR environment in order for that experience to be representative of the medium? VR experiences have a lot in common with video games. Both have virtual environments that the player inhabits and limits on which courses of action can be pursued. Still, there are limitations to what a game designer can allow a player do to. Often, the intended story that the designer is trying to communicate limits what actions players can perform. This is no fatal flaw of the game experience, as the limited agency isn't perceived as immersion-breaking unless it inhibits the narrative coherency or flow of the game. For instance, if you are playing a game in which you are the hero that is supposed to save the world and the end is near, immersion could be broken by the fact that due to a lack of an in-game timer, you can spend years (fictionally) doing chores and other fun activities, whilst ignoring the main quest. But the fact that you cannot build cars in the same game does not detract, *ceteris paribus*, from narrative immersion, although it could be counted as a downside of representative immersion. The transcendental argument simply presupposes that maximizing representative and participatory immersion is all there is to the agentive aspect of VR experiences, and that aspect is key to developing empathy. That would be mistaken; for a narrow focus on representative and participatory immersion leads to an incoherent narrative, which in turn, leaves too little room for developing mentalizing skills.

Conclusion

VR experiences could, in principle, be just as good at enhancing empathy. Currently, VR has not realized that potential. The assessment relies on empirical evidence from studies on empathy enhancement in VR environments. I have provided a how-possibly explanation for the data, while also highlighting the complex interaction between the different types of immersion when one engages in a game of make-believe with a VR environment. As I argued, VR experiences focus too much on participatory immersion through embodiment, sacrificing both representative immersion and narrative immersion. Representative immersion is undermined through a limited virtual environment that lacks promptings for mental acts that would foster moral understanding and mentalizing. Narrative immersion is undermined by an over-reliance on explicit narration and on creating a tight experimental setting which manages to only create a second-hand transmission of the relevant moral knowledge. This, in turn, leads the participant to interact in a doxastically-mediated way with their imaginative projects. Instead, I argued that the same factors that foster developments in mentalizing in other media are also at play in VR experiences: narrative coherency, affective engagement, and prompting moral inferences and inquiries. Theorizing them will be key in analyses on how future VR technology fares when compared to other media.

References

Arora, G., & Milk, C. (2015). *Clouds of sidra*. Retrieved from https://www.with.in/watch/clouds-over-sidra/

Balcerak Jackson, Magdalena & Balcerak Jackson, Brendan (2024). Immersive Experience and Virtual Reality. *Philosophy and Technology* 37 (1):1-24

Bloom, P. (2017a). *Against empathy: The case for rational compassion.* Random House.

Bloom, P. (2017b, February 3). It's Ridiculous to Use Virtual Reality to Empathize with Refugees. *The Atlantic*. Retrieved April 14, 2022, from https://www.theatlantic.com/technology/archive/2017/02/virtual-reality-wont-make-you-more-empathetic/515511/.

Chalmers, D. J. (2022). *Reality+: Virtual Worlds and the Problems of Philosophy*. Penguin UK.

Chasid, Alon (2021). Imaginative immersion, regulation, and doxastic mediation. *Synthese* 199 (3- 4): 1-43

Cogburn, C., Ogle, E., Bailenson, J., Asher, T., & Nichols, T. (2018). *1000 cut journey*. Retrieved from https://vhil.stanford.edu/1000cut/

Gallagher, S. (2020). *Action and Interaction*. Oxford, United Kingdom: Oxford University Press.

Hills, Alison (2020). Moral Testimony: Transmission Versus Propagation. *Philosophy and Phenomenological Research* 101 (2):399-414.

Lara, Francisco & Rueda, Jon. (2021). Virtual Reality Not for "Being Someone" but for "Being in Someone Else's Shoes": Avoiding Misconceptions in Empathy Enhancement. *Frontiers in Psychology*. 10.3389/fpsyg.2021.741516

Martingano, A. J., Hererra, F., & Konrath, S. (2021). Virtual Reality Improves Emotional but Not Cognitive Empathy: A Meta-Analysis. *Technology, Mind, and Behavior*, 2(1). https://doi.org/10.1037/tmb0000034

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450. https://doi.org/10.2307/2183914

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

Persson, I., & Savulescu, J. (2018). The moral importance of reflective empathy. *Neuroethics*, 11(2), 183–193. https://doi.org/10.1007/s12152-017-9350-7

Rijnders, R. J. P., Terburg, D., Bos, P. A., Kempes, M. M., & Honk, J. van. (2021). Unzipping empathy in psychopathy: empathy and facial affect processing in psychopaths. *Neuroscience & Biobehavioral Reviews*, 131, 1116-1126. doi:10.1016/j.neubiorev.2021.10.020

Rueda, J., & Lara, F. (2020). Virtual reality and empathy enhancement: Ethical aspects. *Frontiers in Robotics and AI*, 160. https://doi.org/10.3389/frobt.2020.506984

Sora-Domenjó, Carles (2022). Disrupting the "empathy machine": The power and perils of virtual reality in addressing social issues. *Frontiers in Psychology* 13

Walton, Kendall (1990). *Memesis As Make-Believe*. Harvard University Press.

Walton, Kendall L. (2015). In *Other Shoes: Music, Metaphor, Empathy, Existence.* New York: Oxford University Press.

Zahavi, D. (2014). Empathy and Other-Directed Intentionality. *Topoi* 33 (1):129-142.

Zahiu, A., Mihailov, E., Earp, B. D., Francis, K. B., & Savulescu, J. (2023). Empathy training through virtual reality: moral enhancement with the freedom to fall? *Ethics and Information Technology*, 25(4). https://doi.org/10.1007/s10676-023-09723-9

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, 140(6), 1608. https://doi.org/10.1037/a0037679

Zaki, J., & Ochsner, K. (2016). Empathy. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of Emotions*, Fourth Edition (pp. 871–885). Guilford Publications.