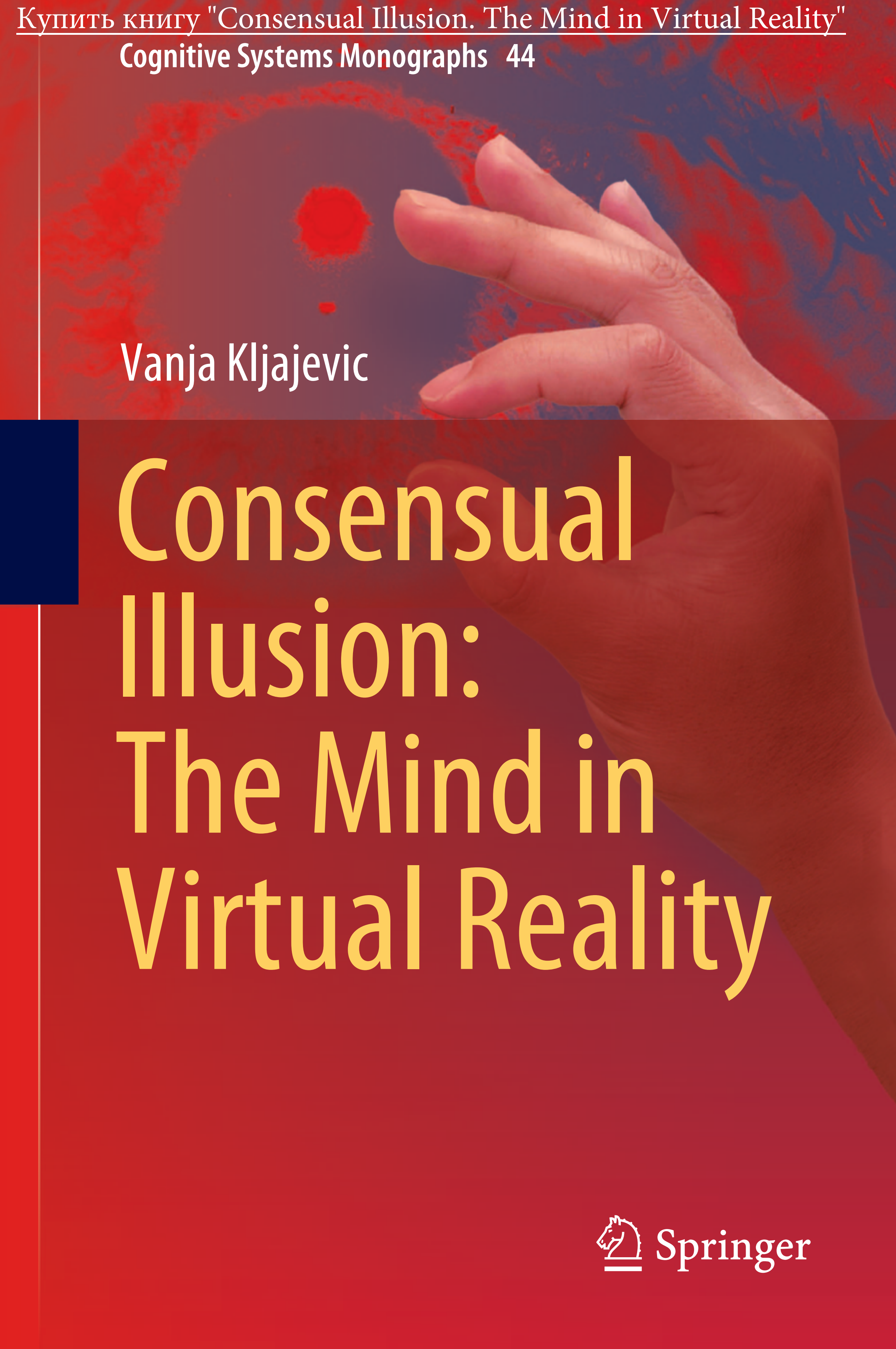

Cognitive Systems Monographs 44

Vanja Kljajevic

Consensual Illusion: The Mind in Virtual Reality

Springer

Chapter 1
A New Kind of Extension

The world is not what I think, but what I live through
(Merleau-Ponty 1958, p. xviii).

1.1 From Real to Virtually Real

The notion that virtual reality is not experienced as another, radically different type of reality, but as an extension of what is already there as physically real, is based on the assumption that there is a reality-virtuality continuum, with the two being positioned at the far ends of the continuum and separated by mixed reality (Milgram et al. 1995). Unlike physically real environments, which by definition remain free of computer-generated interventions, virtual reality environments are on the opposite end of the continuum, being completely computer-generated. Along the continuum we find mixed environments, more specifically augmented reality, in which most of the images are images of real objects, and augmented virtuality, in which most of the images are computer-generated. This concept of display technologies is based on the notion that presence, interactivity and reproduction fidelity are key factors that determine where on the continuum lays a specific application[1] (Gutiérrez et al. 2008).

The term *la realite virtuelle* has been introduced more than 80 years ago by Antonin Artaud in a work related to theatre (Artaud 1938). Since then this loose term has been used to refer to alternative ways of generating experiences of the world, not only by means of theatre, but also by means of literature, film, and other forms of art (Rapolyi 2001). It has been widely popularized in the 1980s, after Jaron Lanier used this term in the context of a new technology aiming to induce

[1]Readers are referred to Gutiérrez et al. (2008) for interesting examples of applications along the reality-virtuality continuum, such as virtualization of the famous Terracotta Soldiers discovered in China in the 1970s, fully immersive virtual tours of archeological sites, and virtualization of ancient Greek theaters, among others, together with an overview of related technical and technological issues, such as the real time rendering of the virtual crowds inside the projected space or interactive sound simulation.

multisensory, interactive, immersive experiences in computer-generated environments (Lanier 2017). Even in this more specific context, the term *virtual reality* remains ambiguous, referring both to virtual reality environments and virtual reality technology. There are additional meanings that have been ascribed to this term, for instance simulation, interaction, artificiality, immersion, telepresence, full body immersion, and networked communication (Heim 1995). However, none of them on its own captures the nature of virtual reality.

Any attempt to define virtual reality must first clarify the meaning of *virtual*.[2] For instance, the *virtual* in virtual reality means that something virtual is real "effectively but not formally", so that

> when a computer graphic makes an entity present to us so effectively that we might just as well have it before us, then the graphic becomes virtually that entity. The graphic then provides a virtual reality (Heim et al. 1995, p. 265).

In addition, *virtual* in *virtual reality* indicates that in a virtual reality system only the model is virtual, and it is so in the sense in which human culture and knowledge are:

> The user really perceives a particular state of the virtual environment. The computer has been actualizing it for the user. Computer systems enable the process of virtualization of a three-dimensional and multi-sensorial model and the process of its actualization. It makes the user feel that this experience is familiar enough to qualify it as reality. (d'Huart 2001, p. 132)

Although there is currently no universally accepted definition of virtual reality, there is a growing understanding across disciplines that virtual environments are computer-generated, interactive and immersive (Burdea and Coiffet 1994; Slater 2009; Gutiérrez et al. 2008; Chalmers 2017). Virtual reality environments differ from mixed environments, which are also interactive and immersive, but only partly computer-generated.

Another controversial term related to virtual reality is *virtual world*. In the context of computer-generated, interactive and immersive virtual reality environments, a

[2]Briefly, the word *virtual* originates from the old Latin *virtus*, which refers to certain powers of human beings. The overtones of power are not present in the Christian meaning of *virtue*, which was "possessing certain values". A fifteenth century definition of *virtual* determines its meaning as being something in effect, essence, or potentiality, but not in actual name or form (Chalmers 2017). In contemporary English, this adjective is generally used to indicate that something is almost as the entity it describes, but not as strictly defined.

The word virtual has more specific meanings in different scientific disciplines. For instance, in computer science it may refer to a software creation that appears as if it exists, although it does not physically exist as such (e.g. *virtual images*) or to something that is accessed or stored on a computer (e.g. *a virtual library*). Lexico (2019) provides the following definitions of *virtual* in some other disciplines: in optics, *virtual* may refer to the "points at which rays would meet if produced backwards", in mechanics to "infinitesimal displacements of a point in a system", while in physics it denotes "particles or interactions with extremely short lifetimes and (owing to the uncertainty principle) indefinitely great energies, postulated as intermediates in some processes". Due to contemporary fascination with virtual reality technologies, the word *virtual* in everyday use is becoming increasingly associated with the meaning found in computer science, i.e. referring to something that is being simulated, being online, or being present not physically but instead in a computer-generated environment.

virtual world is the world generated by using virtual reality technology. However, as the early use of the term *virtual reality* implies (Artaud 1938), virtual worlds are created also by art, which opens the question of whether there are any fundamental differences between these types of virtual worlds. One difference is related to the degree to which these worlds can be shared with others (Chalmers 2017). Virtual worlds created by art, for example when reading a book, watching a movie, or observing a painting, rely on a degree of imagination and thus they cannot be fully shared with others. On the other hand, computer-generated virtual worlds are much more specified and in principle easier to share. There are, of course, differences among virtual environments depending on fidelity. For example, environments with more spatial cues result in more detailed spatial situation models, which increases participants' motivation to participate in such environments and their attention to events, objects and characters they encounter there (Hartmann et al. 2015). Apparently, environments with impoverished spatial cues require a higher degree of participants' visuo-spatial imagery skills to construct stable spatial situational models.

Furthermore, virtual reality applications have been generally divided into those pertaining to realistic virtual worlds and those pertaining to magical virtual worlds (Mania and Chalmers 2001). The former type of environments are simulations of the real world, including for instance flight simulators and other training devices, whereas the latter are environments that defy the laws of physics (e.g. lack of gravity, teleporting to remote objects) and afford superhuman features (e.g. having X-ray vision, the ability to walk through walls, flying by own volition, psychokinesis). Since the early days of virtual reality, it has been emphasized that one advantage of virtual compared to real environments is that the laws of physics that hold for the real world do not necessarily have to hold for a virtual environment. As Sutherland (1965) pointed out, there is no reason why virtual reality has to obey the rules of the physical reality: the true potential of virtual reality is precisely in not having to obey the physical reality laws.

Philosophers who study virtual worlds typically ascribe them a fictional character, assuming that these worlds do not exist in reality. Put differently, such worlds are fictional and so are virtual objects that we encounter there as well as virtual events that take place in such worlds. This view is known as *virtual fictionalism* (Chalmers 2017). Other scholars deny that virtual worlds are fictional, while still ascribing them the status of being unreal, in the same way in which dream worlds are not real. Yet others argue against virtual fictionalism and for *virtual realism*. Briefly, virtual realism is a view that virtual objects really exist, i.e. as digital objects on a computer server, that virtual properties of virtual objects are real, too; virtual events in virtual reality really take place, our experiences in virtual reality are non-illusory and they are as valuable as non-virtual experiences. Virtual worlds are part of the real world via existing on real computers, but that does not make them less real or less valuable (Chalmers 2017).

While philosophers argue about it, for most laymen the physical world exists independently of beings that cognitively represent it.[3] In contrast, virtual reality worlds, being computer-generated, presuppose existence of the physical world and require a thinking, representing mind. Since virtual reality offers a space in which to act, it is a tool that generates the feeling of extension of the physical world. Importantly, even though a computer-generated space remains a virtual environment, the participant's subjective experience of being and acting in that space, the feelings and thoughts generated there, are not fictive, but real.

1.2 Extensions

The notion of extension as an instrument of perception and cognition is certainly not new. For instance, Merleau-Ponty (1958) argued that a stick in the hand of blind man becomes an instrument instead of object of perception, and as such it is an extension of his body. Telescopes and microscopes expand our view of reality so that we can see distant galaxies and observe microorganisms for which we cannot know to exist when we look by the naked eye. Due to these extensions, our senses can reach further.

Anthropologists use the concept of extension to refer to improvements or specialization of human mental and physical functions by means of external tools:

> The computer is an extension of part of the brain, the telephone extends the voice, the wheel extend the legs and feet. Language extends experience in time and space while writing extends language. (Hall 1969, p. 3)

Hall argues that these extensions of human organism have been elaborated to such an extent that they have taken over, "replacing nature" and creating a new dimension, the cultural dimension. Crucially, the new dimension allows us to create "the total world" in which we live and thus to determine who we will be. On this view, humans and their extensions, from cities to technology to language, are extensively interrelated into one system. The extensions we create are in turn shaping who we are. Given this unity, the argument goes, it would be a mistake to consider extensions on their own, i.e. outside of this tightly interconnected relationship with humans, as it would be also a mistake not to pay enough attention to the extensions we create when considering human beings.

Besides Hall, other anthropologists[4] in the 1950s and 1960s argued that human evolution has shifted from the human body to various extensions and that this shift has immensely accelerated the evolutionary process. In other words, using and inventing tools has changed the human nature. Marshall McLuhan had recognized the potential of media as technological prosthesis to reconfigure and change the human nature,

[3]We will not delve into philosophical arguments concerning the nature of reality or deal with intriguing interpretations coming for instance from physics, such as the impact of quantum theory on our understanding of physical reality (Penrose 1989), because such topics would lead us too far from the goals of the present book.

[4]See, for instance, La Barre (1954).

while Clifford Geertz, another anthropologist, argued around the same time for an externalist philosophy of mind, insisting that thinking is not restricted only to the processes that take place in the head but extends to culture (Geertz 1966; Boden 2006). Echoing such ideas, one contemporary cognitive science view posits that mind is "less and less in the head" (Clark 2003, p. 3) and that cognition relies on the body and its interactions with the environment (Barsalou 2008, 2010).

Another interesting concept of extension comes from biology. In 1982 the evolutionary biologist Richard Dawkins introduced the concept of extended phenotype. The phenotypic effects of a gene are the effects that the gene has on the organism in which it sits; an extended phenotypic effect of a gene is the effect it has on the world outside that organism. The idea is that the tools by which the gene propels itself into the next generation "may reach outside the individual body wall", with examples including artefacts such as beaver dams, bird nests and caddis houses (Dawkins 1989, p. 238). Furthermore, extended phenotypic effects of genes in one organism may be found not only on inanimate objects in the world (e.g. stone in the case of caddis houses), but also on the body of another organism, as in the case of snails with extra thick shells, where the thickening is due to the presence of a parasite, in which case it is appropriate to think that the parasite's genes are influencing snails' bodies. The genes can be close to their extended phenotypic effects (e.g. parasites live inside their hosts) or they can act at a distance (a beaver dam is an example of a gene acting at a distance). As Dawkins (1989) puts it succinctly, "It is as if the genes reached outside their 'own' body and manipulated the world outside" (p. 242). Similar to Hall's idea of new dimension, culture can be viewed as Dawkins's extended phenotype in the sense that it continues along the path of biological evolution.

More recently cognitive scientists have recognized the impact of bodily interactions with the world on the evolution of cognition (Thelen 2000), arguing for the concept of extended cognition. Some suggested that extended cognition is "a core cognitive process, not an add-on extra" (Clark and Chalmers 1998) and that "a process is not cognitive simply because it happens in a brain" (Hollan et al. 2000, p. 175). They argue that our minds/brains and bodies together with tool-based extensions form extended thinking systems:

> It is our natural proclivity for tool-based extension, and profound and repeated self-transformation, that explains how we humans can be so very special while at the same time being not so very different, biologically speaking, from the other animals with whom we share both the planet and our genes. What makes us distinctively human is our capacity to continually restructure and rebuild our own mental circuitry, courtesy of an empowering web of culture, education, technology, and artifacts. (Clark 2003, p. 10)

On this view, the unity between the brain, body and technology expands and alters our psychological processes and our sense of who we are. For instance, Clark argues that we are becoming "human-technology symbionts: thinking and reasoning systems whose minds and selves are spread across biological brain and nonbiological circuitry" (Clark 2003, p. 3–4). Regardless of whether we agree with this notion or not, human mind and extensions we invent coevolve.

Further following the development of the idea that tool-based extension of human mental and physical functionality has accelerated the evolutionary process, we arrive at the concept of posthuman. As the concept of posthuman developed, the posthuman future of humanity has been envisioned differently, from post-biological future, in which intelligent machine outperforms human and takes over, to the concept of posthuman as a symbiotic union between human and intelligent machine.

The concept of biological posthuman is related to human engineering of biological cognition via genetic editing. According to one view, such engineering is ideally performed through iterated embryo selection (Bostrom 2014). The goal is to develop superintelligence, i.e. minds that perform significantly better than the best current human minds across multiple cognitive domains.[5] For example, developing new cognitive modules that are comparable in advantage to the way in which language sets apart humans from other species would mean becoming superintelligent.

The concept of posthuman as a symbiotic union between human and intelligent machine assumes that the two components are indistinguishable (Hayles 1996, 1999). Some of the additional assumptions of the posthuman view in this sense are that information pattern has primacy over bodily instantiation, where the body is a "historical accident" and consciousness has a less prominent role than in the traditional Western thought (Hayles 1999, p. 2–3).

Virtual reality technology is an example of tool that allows us to experience having abilities that surpass those characteristic to human beings, a different sense of corporeality, and more generally a different sense of self. However, this does not mean that the concept of posthuman is warranted in this context. It would be misleading to assume that researchers generally agree that the human–machine symbiosis or the way in which virtual reality intervenes in our subjectivity represents a radical break from the past. Kendrick (1996) for example argues that the human–machine symbiosis, or "the technological real", relies on reconstruction of subjectivity by the very technologies that it is supposed to manipulate, and that this type of intervention is not unique to virtual reality. Similarly, Clark (2003) convincingly argues that the biotechnological unions are not more radical than the changes that humanity has encountered in the past, for instance with the advent of timekeeping or the use of text. He remarks that even the early proponents of biotechnology type of symbiosis, including the originators of the term "cyborg", Clynes and Kline, were against the notion that the intervention of new technologies into our subjectivity might transform us into posthuman. New technologies would extend human capacities, for instance by extending the capabilities of the human physical body, but they do not render us posthuman, not because they are not transformative, but because it is in our nature to routinely incorporate such physical and cognitive extensions. It is therefore unnecessary, Clark (2003) claims, to evoke any post-human future, defining ourselves in contrast with, instead of as part of, the world in which we live. Thus, instead of

[5] Other paths to superintelligence have also been envisioned (e.g. artificial intelligence, whole brain emulation, direct brain-computer interfaces, enhancement of networks linking individual minds into collective superintelligence) as well as different types: (i) a speed superintelligence is an intellect that is "just like a human mind, but faster"; (ii) collective superintelligence is a system that includes numerous smaller intellects whose overall performance across general domains outperforms any current cognitive system; and (iii) quality superintelligence is a system that is at least as fast as and qualitatively much smarter than a human mind (Bostrom 2014, p. 53).

arguing that the ways in which new technologies including virtual reality intervene in our subjectivity make us posthuman, a more appropriate view would be to interpret the effects of various tools or extensions within the context of technological progress and our cultural and cognitive adaptation.

Since its inception, virtual reality has fueled debates on what it is and what it should be. These debates reflect distinct theories of reality, distinct philosophies of representation and communication, and differing views regarding how best to benefit from virtual reality. Furthermore, virtual reality challenges the way we think about the relationship between the mind and body. It fools the mind into believing that the participant is in an alternative world that is sufficiently believable to be real, although physically it does not exist as such (Ellis 1991; Vince 1998; Brooks 1999; Steinicke 2016). “Leaving” the body in one environment while inhabiting this other environment may introduce the sense of disembodiment. The feeling of disembodiment is expected to arise in situations when digital immersion is not accompanied by appropriate bodily and environmental feedback. In fact, the sense of disembodiment is relatively common in virtual environments that use head mounted display-based virtual reality systems, because they rarely display a rendering of the participant’s real body. It has been demonstrated that a presence of an avatar as participant’s representation in a virtual environment has beneficial effects on the sense of presence and task performance, and that presence of a self-avatar may affect interaction in shared virtual environments, benefiting cooperative tasks and increasing the level of trust among the collaborators, as well as the sense of presence and perceptual judgments (Pan and Steedman 2017). But as emphasized by Hayles (1999), the point is not in “leaving the body behind”, but in “extending embodied awareness” in ways that would not be possible without the use of specific technology (p. 291).

Virtual reality allows seeing, hearing, touching and manipulating objects in a computer-generated space as if they were physically present. This affects not only participants’ sense of body schema (Sect. 2.3.1), but also their sense of personal location and agency, raising questions such as *Where am I?* (Dennett 2000; Sanford 2000), *Am I where my body is? Am I where the action is?* These questions are relevant for virtual reality because they pertain to one of key features of virtual reality—*presence:* the feeling of being present elsewhere, meaning in the space generated by the virtual reality system instead at the location of the physical body. Virtual reality challenges the notion that our sense of presence is bound to our physical selves, because it allows the participant to temporarily incorporate nonbiological components in his/her body schema and perform actions that may surpass the capacities of real human body or defy the laws of physics. It extends the participant’s sense of corporeality beyond the borders of the physical body, and by introducing the concept of virtual body it challenges the idea that one agent corresponds to one body.

Importantly, even when actions performed in a virtual reality environment are beyond the bounds of the physical reality, there is a degree of awareness that the world of enactment is enabled by a specific technology and embedded in the physical world via a specific discourse (Slater 2009). The ability to retain this awareness is considered to characterize all but naïve users of virtual reality (Chalmers 2017). An important implication of this view is that presence allows a degree of awareness about the real world, and yet the virtual space has the power to induce the feeling that the participant is there, in the virtual environment.

1.3 *Being* … Where?

Presence, which is short for the term *telepresence*, is typically defined as a subjective feeling of *being there*, meaning being virtually present in an alternative environment, at a location different from the actual location of the physical body. The concept of presence was introduced by André Bazin in 1951 in the context of film experience and in the 1970s it was extended to communication phenomena mediated by technology (Lombard and Jones 2015).

Due to Minsky's (1980) seminal paper on telepresence, followed by the establishment of the MIT journal *Presence: Teleoperators and Virtual Environments* in 1992, presence has emerged as a concept crucial for technology-mediated virtual environments. Namely, Minsky's term telepresence referred to remote presence mediated by technology:

> To convey the idea of these remote control tools, scientists often use the words 'teleoperator' or 'telefactor'. I prefer to call this 'telepresence', a name suggested by … Patrick Gunkel. Telepresence emphasizes the importance of high-quality sensory feedback and suggests future instruments that will feel and work so much like our own hands that we won't notice any significant difference. …The biggest challenge to developing telepresence is achieving the sense of 'being there'. Can telepresence be a true substitute for the real thing? Will we be able to couple our artificial devices naturally and comfortably to work together with the sensory mechanisms of human organisms? (Minsky 1980)

While most researchers agree that presence is one of the most noticeable psychophysical effects of immersive virtual reality (Waltemate et al. 2018), there is little agreement on how to best define this concept. In fact, there are currently many conceptualizations and definitions of presence, and many different terms are used to refer to the feeling of *being there,* such as telepresence, co-presence, spatial and social presence, virtual, immersive, perceived, subjective, and so on. For example, presence has been defined as the sense of being located at the place depicted by the virtual displays (Sheridan 1992), as an all-or-none psychological phenomenon (Slater 2009) associated with the illusion of being present at a location different from the actual location of the physical body (place illusion), which is "a 'response' to a system of a certain level of immersion" (Slater 2003). Others define presence as "the feeling of being located in a perceptible external world around the self" (Waterworth et al. 2015), as "the experience of being engaged by the representations of a virtual world" (Jacobson 2002), and as the "perceptual illusion of nonmediation" (Lombard and Ditton 1997; Lombard and Jones 2015).

Slater (2002) has noted that defining presence in terms of the feeling of *being there* is a category mistake and that the sense of being elsewhere is just one of many contributors to presence. Following this intuition, Riva (2009) defined presence in terms of agency and control: "subjects are 'present' if they are able to enact in an external world their intentions" (p. 159). This definition is consistent with the observation that our sense of where we are located largely depends on the sense of an "action-space" (Clark 2003, p. 94). Similarly, Wirth et al.'s (2007) model of presence formation requires two steps: self-localization in a mediated environment and the perception of possibilities for action in that environment. Given that presence is tied to the notion of space by definition, it is not surprising that it is sometimes defined in terms of what can be done in that space. Anthropologists in the 1960s emphasized this dynamic aspect of our perception of space (e.g. Hall 1969).

The variety of definitions of presence has caused a great deal of confusion regarding what constitutes presence, not only among designers and consumers, but also among scholars, impeding progress in the field (Slater 2003). Thus, tidying up the definitions of presence would be helpful. Since different researchers use the term *presence* differently, a unifying theory of presence is currently not possible (Waterworth et al. 2015). The International Society for Presence Research (2000) defines presence as

> a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by/or filtered through human-made technology, part or all of the individuals' perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at *some level* and *to some degree*, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience (www.ispr.info/about-presence-2/about-presence/).

For instance, using a sophisticated flight simulator we may, at least for a short while, be completely unaware that our experience of flying an aircraft is actually being mediated by the technology.

> When you are present your perceptual, vestibular, proprioceptive, and autonomic nervous systems are activated in a way similar to that of real life in similar situations. Even though you cognitively know that you are not in the real life situation, you will tend to behave as if you were, and have similar thoughts…. (Slater 2003, p. 2)

Thus, when the participant feels as being present in an immersive virtual environment, certain knowledge that the environment is mediated is always there, if not in the person's full awareness, then on its border, ready to enter the focus. It is therefore important to determine the factors that channel the processes related to this awareness and how they lead to shifts between the participant's feeling of being in and being out of the virtual space. We address these issues in the following sections.

1.3.1 Breaks in Presence

When fully immersed in a virtual environment, a participant is subjected to two streams of sensory data, one stream coming from the physical world and the other from the virtual world displayed by the virtual reality system (Slater et al. 2003). If for some reason the participant stops responding to the data from the virtual environment and instead responds to the sensory data from the physical world, a break in presence will follow. Recording the number and duration of such breaks can help improve presence in virtual environments. Such a method of analysis is much preferred to subjective reports and presence questionnaires. The latter methods of assessment have been critiqued on methodological grounds (Mania and Chalmers 2001; Slater et al. 2003), for instance as being unreliable, because participants report subjective, post-hoc impressions about their virtual reality experiences some of which may not be accurate and at the same time they may forget to report important indicators of transition between the environments. Physiological responses signaling such transitions are therefore considered a better basis for measuring presence than introspection-based methods.

Slater (2002) outlines a way of thinking about presence that, instead of relying on an account of experience based on introspection, postulates that presence is a perceptual mechanism that answers the question *Where am I?* That is, it selects between the hypotheses that a person is in a virtual environment vs. that he/she is in the real world. It is this switching between the alternative environments that makes presence such an interesting research topic: if we receive signals from only one environment, we are present in that environment; but when there are competing signals from different environments, how do we choose the signals based on which to act?

As an example, a participant who is walking in a virtual room with a precipice in its center—while actually standing in a CAVE-like system or wearing a head-mounted display and feeling the cables, temperature and the physical signals of the actual place—will experience presence if the virtual environment sensory signals systematically override other sensory signals. In fact, this can be achieved using minimal cues:

> At some deep level, our minds do not understand "virtual reality". Hence, only minimal cues are necessary for our presence selection mechanism to respond to a virtual pit, even though we know "really" that there is not one there. The sixth sense is this process of seeing what we expect to see, and it doesn't take much for a virtual reality to convince us: we respond to events in the virtual world much as we would to similar events in the physical world. (Slater 2002, p. 435)

Even though participants have an abstract knowledge that they really are located within a CAVE space and not in a virtual room with a precipice, "visual perception overrides this knowledge and the bodily system reacts as if they were in a virtual room: heart rate rises, locomotion is carefully judged, and the subject reports symptoms of anxiety" (Slater 2002, p. 436).

Slater suggests that the presence selection mechanism is grounded in the previous work on perception that assumes that perception consists of selection between hypotheses. For instance, the explanation goes, choosing between the hypotheses in perceiving the famous Kanizsa triangle typically leads to an optical illusion, i.e. seeing the edges of the white upright triangle between the circles even though such edges are not there. In addition, Slater credits Norton and Stark's (1971) scanpath theory of eye movements for his notion of presence selection mechanism, more specifically the top-down cognitive model of visual perception that this theory implicates. According to this model, what we perceive in the world depends on what is in our mind's eye. When observing ambiguous figures in which we can recognize two objects (e.g. a duck and a rabbit), our eye movements change according to whether we decide that the figure represents a duck or a rabbit. On this view, a top-down cognitive mechanism, together with the image, determines our eye movements, which suggests that eye movements and perception are not fixed by what is represented.

Note, however, that this explanation contains the assumption that cognitive states penetrate visual perception, which is debatable. In fact, the question of whether perception is cognitively penetrable is at the heart of one of the most interesting debates in cognitive science.

1.3.2 Cognitive Penetrability of Visual Perception and Presence

The issue whether vision is continuous with cognition, i.e. cognitively penetrable, or discontinuous, distinctive, and independent from it has been debated for a long time.[6] One formulation of cognitive penetrability is as follows:

> …if a system is cognitively penetrable then the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows. (Pylyshyn 1999, p. 343)

Those who argue that vision is continuous with cognition postulate that vision processes are affected by higher cognitive processes. Those who argue against this view postulate encapsulation or cognitive impenetrability of certain visual processes, such as early vision processes.

One recent approach to visual perception suggests that vision is impervious to cognitive penetration and that whatever influence cognition may have on sensory stimuli processing, it may have to do not with cognitive states (beliefs, desires, emotions, motivations, intentions, linguistic representations), but with other factors, such as attention:

> Perhaps most prominently, shifting patterns of *attention* can change what we see. Selective attention is obviously closely linked to perception – often serving as a gateway to conscious awareness in the first place, such that we may completely fail to see what we do not attend to (as in intentional blindness…). Moreover, attention – which is often linked to a 'spotlight' or 'zoom lens' … – can sometimes literally highlight or enhance attended objects, making them appear (relative to unattended objects) clearer … and more finely detailed….. (Firestone and Scholl 2016, p. 13)

According to this view, selectively attending to a different feature of an object is similar to changing what we see when moving our eyes. In other words, what

[6]See Pylyshyn (1999) for a brief review of the debate, including the views held by von Helmholtz, Bruner and the New Look in Perception, among others.

changes is the input to mechanism of visual perception, whereas visual processing itself remains fixed, that is, unaffected by cognitive states. For instance, when looking at an ambiguous image, its disambiguation does not require penetration of perception by any top-down state (intention, emotion, desire); instead, voluntarily switching between two interpretations (e.g. duck/rabbit), although it depends on peripheral attention, changes only the input to visual processing:

> … though one may indeed be 'changing one's assumptions' when the figure switches, this is not actually triggering the switches. Instead, the mechanism is that different image regions are selectively processed over others, because such regions are attended differently in relatively peripheral way. (Firestone and Scholl 2016, p. 14)

This view predicts that the presence selection mechanism may be not perceptual, but rather an attention-based mechanism. It is compatible with the idea that presence in a virtual environment is an experience "similar to that of physical reality, together with an imperceptible shifting of focus of consciousness to the proximal stimulus located in that other world" (Sas et al. 2004, p. 53). In other words, switching between experiencing alternative environments—the immediate and the mediated—is not due to a cognitive state penetrating perception, but it is rather based on the change in a visual input due to a shift in attention.

However, despite the evidence supporting this view regarding vision (e.g. Gur 2016) and other domains of perception, such as speech perception processes (e.g. Cutler and Norris 2016), the view that perception is not penetrable by cognition has been challenged on several grounds. For instance, the model of vision inherent to this view consists of three modules: a sensory/input module, a perception module, and a cognition module (Firestone and Scholl 2016). Yet the neural evidence on the organization of the visual brain does not seem to support such neatly differentiated and entirely encapsulated modules, indicating instead that the primary visual cortex (V1) is involved not only in an early stage of vision but that it is actually involved all the way through, supporting feedforward and feedback activity of the network; that attention effects do not proceed from input to perception but rather they cause changes across the visual hierarchy; and that overall interactions among the brain areas involved in vision are so extensive and complex that it is "impossible to neatly separate sensation and perception":

> Rather than simply gating or enhancing input, much of the neural hardware responsible for vision flexibly changes its function in complex ways depending on the goals of the observer. (Beck and Clevenger 2016, p. 20)

While the original notion of encapsulation proposed by Fodor (1983) holds that "at least some of the background information at the subject's disposal is inaccessible to at least some of his perceptual mechanisms" (p. 66), Firestone and Scholl's (2016) notion of encapsulation is stricter, because they claim that the most fundamental distinction in cognitive science is "between seeing and thinking", and that "perception proceeds without any direct, unmediated influence from cognition" (p. 17).

The question whether perception and cognition are neatly separable or whether the line between them is blurred, reflecting a continuum in information processing capacities, has been vigorously debated in cognitive science. The question has theoretical and experimental value, as well as clinical relevance because of its role in contemporary approaches to the formation and maintenance of delusions (Ross et al. 2016). Other developments in cognitive neuroscience largely support the view that processing in the human brain is distributed and context-sensitive (Hackel et al. 2016), that the brain is a predictive organ, and that information processing is predictive rather than merely integrative (Federmeier 2007; Friston 2012; Hoffmann and Falkenstein 2012; Van Petten and Luka 2012). This evidence comes from research including vision, attention, motor control, motor imagery, action understanding, music, language processing, emotional processing, executive functions, and theory of mind. It suggests that the brain uses the incoming stimuli, which are noisy and partial, to create highly structured models for making predictions, which are then compared against the input.

Regardless, it is worth considering Firestone and Scholl's (2016) approach to visual perception and their contribution to the debate on cognitive penetrability of perception, because their approach opens a path to the explanation of breaks in presence that emphasizes the role of attention in shifting between feeling present in the real and a virtual environment. Specifically, assuming that digitally generated environments may compete for processing resources with the participants' immediate environments, an attentional model of presence was developed. According to the model, attention is the most critical component of presence, because it determines whether the participants' sensory perception is directed to mediated stimuli or to cues from the immediate environment (Draper et al. 1998; Wirth et al. 2007). While it is obvious that different cognitive resources play a role in our perception of the environment we are in, it has been pointed out that attention on its own is insufficient to evoke the sense of being located at a place different from the immediate environment and that attentional resources can be exhausted without evoking the sense of presence (Hartmann et al. 2015). However, shifts in attention can explain breaks in presence.

1.3.3 Some Other Factors Affecting Presence

One widely accepted assumption is that participants in virtual environments process the sensory information using the mental models they use in everyday life (Slater and Usoh 1994). Anthropologists have long observed that people from different cultures belong to different sensory worlds (e.g. some cultures rely more on olfaction and touch than others), applying different sets of culturally-molded sensory filters (Hall 1969). Such cross-cultural differences need to be considered in virtual reality design, because a specific sensory input may have different effects on participants who belong to different cultures.

There is also evidence indicating that a cognitive style may influence the sense of presence (Sas et al. 2004). The concept of cognitive style originates from Jung's (1971) theory of psychological types and refers to enduring patterns of cognitive behavior. More specifically, cognitive styles are consistent modes of perceiving, remembering, judging, and problem solving that characterize an individual (Messick 1989). They are tightly linked to affective, temperamental and motivational aspects of personality structure. Implicit in these patterns are people's mental models of the world and their processing biases. Evidence from a virtual reality study suggests that people who are more introvert and sensitive often experience a higher degree of presence than extroverts (Sas et al. 2004). The study suggests that depending on a type of task they want to complete in such environments, introverts may need additional support, and that since for instance they appear to have a diminished level of spatial awareness, their performance on spatial tasks would benefit from additional cues, such as additional landmarks. In contrast, extraverted people do not need such support, because they appear to have increased level of spatial awareness. While more research on the topic is necessary, the findings that cognitive styles affect presence and task performance in virtual reality indicate that design of virtual environments needs to consider the fact that people differ with regard to patterns of cognitive behavior they adopt.

Not only that characteristics of participants' cognitive styles affect their experience of a virtual environment, but personality traits of virtual characters may have that effect as well. A recent study investigated how personality traits in virtual characters affect participants' perception of uncanniness (Sect. 3.1.2). The study reported that the perception of personality traits indicating psychopathy, such as virtual characters' lack of the startle response to a scream, was a strong predictor of uncanniness, whereas perception of other negative personality traits that are not indicative of psychopathy was not (Tinwell et al. 2013). Stein and Ohler (2017) suggest including personality assessment variables as covariates when studying the uncanny valley effects. Thus, personality traits of virtual characters are an important question, because the feeling of uneasiness may affect presence and participants' task performance.

Although individual variation in human cognitive task performance as well as in brain anatomy and plasticity is a well-established fact, it is usually taken for granted that there is a single "typical" human cognitive architecture. This notion has recently been criticized and it was suggested that there may be more variability in the "mapping from brain function to mind" than previously recognized and that human cognitive architecture may "differ substantially across individuals" (Adolphs 2016, p. 1153). Relating this notion to presence, individual differences in susceptibility to a specific sensory modality (vision, touch, sound) affect the degree to which that modality may induce presence. Given the elusive nature of presence, one could argue that there may exist differences in susceptibility to a specific sensory input even within a single participant across different contexts. This once again highlights the importance of consistency of incoming sensory data, which has been emphasized as the primary requirement for presence (Slater 2003). Furthermore, not only differences in the processing of incoming sensory input, but also differences in participants' intentions may lead to differences in experiencing presence across participants in immersive virtual environments (Riva 2009).

Another factor that may differentially influence participants' experience of presence and their task performance in virtual reality is gender. The notion that gender may play an important role in cognitive functions has received much attention in the past and it still continues to provoke debates. The idea that women in general outperform men in certain cognitive tasks, such as verbal learning, whereas men in general outperform women in other tasks, such as math and spatial reasoning, had dominated psychological science in the 1960s and 1970s. Although not so much in the foreground, the debate on cognitive gender differences continues, generating further evidence on female vs. male advantage for cognitive tasks (e.g. Kimura and Clarke 2002). On the other hand, the gender similarity hypothesis posits that males and females are similar in most cognitive abilities, including language and math, with large differences apparently existing in motor abilities, sexual behavior, and aggression (Hyde 2005, 2016). Importantly, the question of whether gender differences exist in experiencing presence and task performance in virtual reality has not been sufficiently addressed so far. Some scholars argue that studies on virtual reality need to control for gender, pointing out differential (females > males) susceptibility to cybersickness and the impact of sex hormones on cognition (Larson et al. 1999; Parsons et al. 2004).

For example, supporting the view on gender differences in spatial processing, an early study on mental rotation and spatial rotation in a semi-immersive environment reported that male participants performed considerably better in the traditional pen and pencil mental rotation task, but that there were no statistically significant gender differences in participants' performance on the virtual reality spatial rotation task (Larson et al. 1999). The results were replicated in a later study by the same group (Parsons et al. 2004). Importantly, the two tasks differed only in one aspect: unlike the paper and pencil version of the test, which required mental rotation, the virtual reality spatial task allowed manual manipulation of the stimuli. Since the two tasks were not equivalent in the two environments, it is difficult to directly compare the results obtained on the tasks, although the study remains informative regarding the gender differences. Coutrot et al. (2018) also found a male advantage in navigation, which was more prominent in a virtual than in a real environment.

Clearly, the impact of gender on presence and task performance in virtual environment is an important issue, and it is disappointing that some recent studies that used sophisticated methods such as functional magnetic resonance imaging (fMRI) for assessment of spatial abilities in virtual reality included only male participants (e.g. Mellet et al. 2010). By not including female participants, such studies miss the opportunity to contribute potentially invaluable evidence regarding this largely neglected factor which may confound cognitive results and lead to a distorted picture on participants' performance in virtual reality and their sense of presence.

1.3.4 Lost and Inverse

Another perspective in defining presence is to determine what it is not. For instance, to negatively define presence, Slater and Usoh (1994) suggest as key feature the absence of the sense of location, "such that 'no presence' is equated with no locality, the sense of where self is as being always in flux" (p. 130). To illustrate the point, they refer to Oliver Sack's patient who lost the ability for present day memory as a case of neurological loss of presence. Briefly, this patient would forget with whom he was talking and what the topic of conversation was, losing the context every few moments. "Imagine", write Slater and Usoh, "a VR system that continuously and randomly changed the environment, so that the human participant could form no stable sense of locality, and no relationship with any object: everything being continually in flux. Such an environment would not be presence inducing" (p. 130). This nicely reiterates the danger of breaks in presence, and the critical roles of flow and continuity of the feeling of being in one location as well as the importance of stability of the relationship between the self and the surrounding objects.

Finally, an experience related to presence is *inverse presence.* Namely, while presence pertains to mediated experiences that involve an illusion of being non-mediated, the concept of inverse presence refers to those experiences that are not mediated and yet they involve an illusion of being mediated (Timmins and Lombard 2005). Experiences of inverse presence include, for instance, the feeling of being within a video game while participating in real world activities. Such experiences involve a degree of disbelief and are usually associated with descriptions such as "I thought I was watching a movie", "I couldn't believe my eyes", and questions such as "Am I really here?"

> The inverse-presence experience allows the person to pretend, at least at some level, that the event is not real and not "really" happening; because the event seems like a mediated experience that is therefore not real, it can serve to distance the person from, and help him or her cope with, the unpleasant reality. (Timmins and Lombard 2005, p. 498)

While one common feature of presence and inverse presence is a distorted picture of the role of medium in the experience, an important difference is that they appear in contexts that are opposite (mediated, not mediated) and involve different cognitive mechanisms. Presence requires suspension of disbelief, while inverse presence requires a degree of disbelief. Presence hinges on the ability to shut out the cues to the external environment, whereas inverse presence requires a sustained effort to focus on a selection of cues from the external environment and enhance them along some personally relevant dimension that regulates the way in which an individual experiences the outside world.

1.4 Physicality Expectations and Pathways to Illusion

Efforts to develop virtual reality environments that elicit presence have focused more on issues related to location than on action, revealing a need for a richer physical experience of immersive virtual environments (Giannopoulos et al. 2011). Researchers often point out to examples such as hitting a wall, picking up an object, or shaking somebody's hand in virtual environments. To be believable, such actions require a degree of physicality. If the expected physical feedback is lacking, participants "feel nothing" and such actions and situations are not believable. Current generally available virtual reality technology does not involve systems with generalized haptics, and if the participant touches an object randomly, they will feel nothing:

> The whole aspect of physicality is typically missing from virtual environment experiences—collisions do not typically result in haptic or even auditory sensations. (Slater 2009, p. 3550)

In an immersive virtual reality study of interpersonal distance in which participants interacted with two types of virtual characters, embodied agents and avatars, the lack of physical contact between virtual characters was experienced as a major disadvantage of using virtual environments (Bailenson et al. 2003). Receiving haptic feedback when jointly manipulating an object in an immersive virtual environment increases the sense of social presence (Oh et al. 2018). In general, improved haptics would ensure a more realistic experience, regardless of whether a virtual environment is intended for training, therapy, or other activities that involve physical contact. Feeling the texture and weight of objects in a virtual environment, or the strength of a grip and trajectory of movements while shaking hands with other virtual characters are equally important as visual and auditory cues (Slater 2009).

In an immersive virtual environment real sensory data are replaced by computer-generated data (Mel Slater 2008), yet certain physicality is necessary. How is this to be achieved?

1.4.1 Virtual Embodiment

Virtual embodiment is a process of identifying with a virtual (computer-generated) body, i.e. an avatar, which is a digital proxy of the participant in a virtual environment. An avatar represents the participant's physical body in a virtual environment so that, although located in a different environment (e.g. a laboratory), the participant acts in the virtual environment by means of controlling his/her avatar from this other location.

Virtual embodiment may also refer to the end point in the process of identifying with a virtual body, whereby the participant experiences illusion of ownership over a virtual body. This gives the concept of avatar a unique role in virtual environments. Avatars are often conceptualized as "virtual replicas" or digital alter egos of the participants' physical self. According to one definition, avatars are "the direct extensions of ourselves into the virtual domain"; they are "the digital representations tightly bound to our embodied self, our self perception and our personality" (Waltemate et al. 2018, p. 2). Their function is to enable interaction between the participant and a virtual environment, other virtual characters, and objects encountered in the virtual world. It is not surprising then that avatar's appearance and behavior cause a range of psychophysical effects in participants they represent as well as in other participants with whose avatars they interact.

Illusion of ownership over a body part or over a whole body has been induced in immersive virtual environments for instance by using synchronized visuo-tactile stimulation, as in rubber hand illusion. This way of mapping of the participant's body schema onto his/her avatar takes place via afferent or sensory signal correspondences (Bailey et al. 2016). In addition, the mapping can be realized via sensorimotor correspondences between the physical body and the virtual body. Thus, virtual reality affords different ways in which body ownership illusion may be induced, involving bottom-up factors, which are related to multisensory integration (visual, motor and tactile stimulation), and top-down factors, which are related to conceptual aspects of the experience, i.e. interpretation of the observed virtual body (e.g. its appearance) (Waltemate et al. 2018).

A concept that plays an important role in understanding of virtual embodiment is embodied or grounded cognition.[7] It assumes that cognition is not limited to what is inside the skull, but is grounded in the body and its relationship to the environment:

> … the cognitive system utilizes the environment and the body as external information structures that complement internal representations. In turn, internal representations have a situated character, implemented via simulations in the brain's modal systems, making them well suited for interfacing with external structures (Barsalou 2010, p. 717).

Applying this notion to virtual environments, one may wonder about how mental representations of objects, events, characters and their interactions in immersive virtual environments are grounded. It emphasizes even more the role of the physical body in virtual reality and the complexity with which this role is projected to a computer-generated space, given that the bodily feedback is mediated and thus limited or sometimes completely lacking in such environments. An additional level of complexity is involved when virtual bodies offer affordances different from those of human body, such as an avatar with a tail (Steptoe et al. 2013) or an avatar with three arms (Won et al. 2015). This opens an intriguing question of how far virtual embodiment can go and still allow mapping of virtual objects onto human body schema.

Steptoe et al. (2013) introduced the concept of "extended humanoid avatars" to refer to avatars with fundamentally human form, but with some additional features. In this specific study, avatars had a movable, long tail, extending 0.5 m beyond each arm. The study involved an immersive virtual CAVE-type environment with 32 participants, whose movements were tracked and whose avatars reflected their movements. One half of participants reported that the avatar's tail moved in a random and asynchronous way relative to their movements, whereas the other half reported that the avatar's body moved in a synchronous way and that they could control it by hip movements. Moreover, the participants who controlled avatar's tail movements experienced anxiety when faced with a virtual threat to the tail. This finding is interesting because it indicates that regardless of the appearance of avatar's body, these participants were able to experience body ownership, confirming once again the flexibility of the human body schema.

---

[7] The term *grounded cognition* is sometimes used interchangeably with the term *embodied cognition*. However, the term *embodied cognition* overemphasizes the role of the body in grounding, whereas the physical and the social environments are equally important grounding mechanisms (Barsalou 2008, 2010).

Avatars are used to visually represent participants in virtual environments. In shared and mixed reality environments, these are interconnected participants. Importantly, virtual reality affords an unprecedented degree of possibilities for extending and radically changing our virtual bodies. In addition to avatars' bodies that considerably differ from the human body matrix and yet are evolutionary plausible, other types of virtual bodies have also been used to investigate body ownership in virtual reality, including bodies with additional limbs or bodies whose structure has no basis in evolution (Bailey et al. 2016). Although virtual, these bodies are still perceptually real (Steptoe et al. 2013). The sense of ownership over the virtual body and the sense of control over the virtual body's actions have been reported for human-like and other types of avatars.

However, regardless of the sense of ownership over the virtual body and the sense of control over its movements, the process of motor prediction differs between a virtual body that has a counterpart in the physical world and a body that does not (Steptoe et al. 2013). Consider as an example an avatar with a tail. Although an intention to move avatar's tail is followed by an efferent signal to the hip muscles which perform the movement and afferent signal from proprioception is sent to the brain, and although the visual signal confirms the hip movement as well as the presence and the movement of the tail, the tail's movement itself cannot be felt. And this is what sets apart the avatar's tail movement experience from the normal body movement experience and places it closer to a phantom limb experience. While in the phantom limb experiences people can sense but cannot see the movement, in the avatar tail experience participants can see but cannot sense the movement. The dominance of the visual feedback in the case of avatar with a tail may attenuate, but it cannot compensate for the lack of feedback on how the tail's movement feels.

The Steptoe et al.'s observation is significant, because it suggests a clear demarcation between virtual embodiment based on human-like and extended humanoid avatars. The difference has implications for design of virtual reality applications for treatment and training as well as entertainment. On a theoretical level, it suggests the importance of studying effects of virtual embodiment. For instance, one relevant question is: *Can a movement that is not felt still be believable?* Apparently, even if it is not felt, as long as the movement can be seen and is synchronized with participant's movements, it is believable. And being believable is what counts most in this context, or what makes virtual reality feel real (Brooks 2003). Research evidence converges in indicating that the first-person perspective, visuo-tactile and visuo-motor synchrony as well as sensory feedback are critical steps in inducing virtual embodiment, and that even just observing a virtual body from the first person perspective leads to virtual embodiment (Slater et al. 2010).

Assuming that participants in immersive virtual environments can learn how to control their virtual bodies, Won et al. (2015) investigated embodiment of an avatar that has been designed with affordances for tasks not available to human body. One question in the study was whether participants would use avatar's body in a way that is optimal to accomplish the task, thereby modifying the body schema, and another question was whether they would benefit from the affordances for task completion