# The consequences of living in a virtual world generated by our brain

Jan Westerhoff
(University of Durham)

## Abstract

Recent discussions in cognitive science and the philosophy of mind have defended a theory according to which we live in a virtual world akin to a computer simulation, generated by our brain. It is argued that our brain creates a model world from a variety of stimuli; this model is perceived as if it was external and perception-independent, even though it is neither of the two. This theory constitutes a radically new approach to thinking about the the mind, the brain, and the world, an approach the peculiar consequences of which have rarely been explored in detail.

It differs from indirect realism as traditionally conceived insofar as the perceiver (the person or the self) is supposed to be part of the simulation as well, an idea that gives rise to an intriguing circularity, since it appears as if the "generating system" that brings about the simulation already has to be an intentional agent in order to produce the intentional agent that is the self. More importantly, I will argue that this theory has the surprising philosophical consequence that we can no longer be realists about the external world that is supposed to supply the stimuli on which the simulation is based. If we live in a brain-based simulation both the notion of a perceiver "in here" as well as that of a perceived world "out there" become untenable.

Virtual world theory (VW theory for short) is a widespread conception of perception that can be found in various contemporary discussions of philosophy, cognitive science, and phenomenology, as well as in the popularization of biological science. According to this theory conscious experience is a type of virtual reality, a virtual world generated by our brain that constitutes a model of the real world. The clearest and most sophisticated contemporary exponent of VW theory is Thomas Metzinger (2003, 2010). Another explicit (if philosophically less acute) proponent is Richard Dawkins who dedicates a substantial amount of space in "Unweaving the Rainbow" (1999) to support the theory that we live in a virtual world. He notes that:

"We move through a virtual world of our own brains' making. Our constructed models of rocks and of trees are a part of the environment in which we animals live, no less than the real rocks and trees that they represent." [Dawkins 1999: 284.]

"You and I, we humans, we mammals, we animals, inhabit a virtual world, constructed from elements that are, at successively higher levels, useful for representing the real world. Of course, we feel as if we are firmly placed in the real world - which is exactly as it should be if our constrained virtual reality software is any good. It is very good, and the only time we notice it at all is on the rare occasions when it gets something wrong. When this happens we experience an illusion or a hallucination [...]" [Dawkins 1999: 275-276.]

Quotations like these can easily be multiplied when examining contemporary literature on the brain and the mind.

VW theory seems to have plenty of empirical support on its side. Seeing an orange on the table in front of us it is clear that the orange does not go inside our head, nor do we go out to the orange by means of some sort of perceptual rays. Rather, in various indirect ways the orange stimulates our nerve-endings by contact with different sensory organs, these then pass the stimuli on to the brain where a perception of the orange including its various visual, olfactory, tactile, and perhaps auditory aspects are put together. This perception is part of the virtual model of the world in which we live our lives. Of course this entire process is hidden from us. It is, as it is sometimes said "transparent", because we see right through it, like a pane of clear glass, to the end result, the virtual orange. We do not have a the possibility of observing our brain putting together the virtual orange from the various stimuli *from the outside*, as we can e.g. for a director producing a film, or a programmer producing a computer simulation.

So far, so familiar. We might very well ask whether VW theory is not just old epistemological wine in new neuroscientific bottles, a somewhat updated version of indirect realism familiar at least since the days of Plato's cave. Standard-issue indirect realism presents the following picture of our cognitive relation to the world:

perceiver -> percept ||veil of perception|| <- object

On the left-hand side are we, the perceiver, on the right-hand side is the object we perceive. Somehow the object gives rise to the percept, something we can be in direct contact with, unlike the object, which is shielded from us through the veil of perception. VW theory differs from this conception of indirect realism through its very different conception of what goes on at the left-hand side of the veil, while it more or less agrees about what happens on the right-hand side.

Rather than postulating the existence of percepts and a perceiver on the left-hand-side of the veil VW argues that the perceiver is no less of a construct than the virtual world he perceives. While the view of reality suggested by indirect realism is reminiscent of a pilot in a flight simulator, VW introduces the notion of a *total flight simulator*:

"The brain is like a *total flight simulator*, a self-modeling airplane that, rather than being flown by a pilot, generates a complex internal image of itself within its own internal flight simulator. The image is transparent and thus cannot be recognized as an image by the system. Operating under the condition of a naive-realistic self-misunderstanding, the system interprets the control element in this image as a nonphysical object: The "pilot" is born into a virtual reality with no opportunity to discover this fact. The pilot is the Ego." [Metzinger 2010: 108].

This construction is intriguing, not least because of its obvious circularity. The "system" here appears as an intentional agent (it "interprets") producing another intentional agent (the Ego). The construction of the self in this example proceeds by appeal to an intentional agent, but this in turn seems to depend on the existence of a self that can bring about this very intentional agency. The question whether this circularity is vicious, or whether this notion of the constructed self in unsatisfactory for other reasons is beyond the scope of the present remarks. But it is clear that the theory described here is very different from indirect realism as we know it. While indirect realism can easily slide towards solipsism ("What if the only real things are on the left-hand side of the veil?") VW is more prone to move into the other direction, towards anti-solipsism, a position that holds that the only real things are objects *other* than oneself. For most varieties of VW theory the perceiver on the left-hand side of the veil cannot be considered as fundamentally real.

On the other hand VW theory in general has no unusual views of what is going on at the right-hand side of the veil. Metzinger (2010: 23) is very clear about the existence of an external (or, as he sometimes calls it, "extradermal") reality. Yet if it is the case (as he points out) that we cannot have "conscious experience of knowing" this reality, a version of transcendental idealism seems to be our best bet. Of course not all proponents of VW theory share this quasi-Kantian picture. Even though he is not very explicit regarding the details, Dawkins' conception of the "real" world behind the simulation sounds considerably more substantial that the Kantian noumenon:

"We are so used to living in our simulated world and it is kept so beautifully in synchrony with the real world that we don't realize it is a simulated world." [Dawkins 1999: 281-282.]

Advocates of VW theory are not very forthcoming with about arguments for their assumptions as to what is going on at the right-hand side of the veil. However, the following passage provides us with some clues:

"Trivially, if an internal representation of the system itself exists, according to the fundamental assumptions of any naturalist theory of mind there also has to exist a physical system which has generated it. I call this the “naturalist variant of the Cartesian cogito.” Pathological or systematically empty self-representata may exist, but their underlying existence assumption will never be false, because some kind of constructing system has to exist. Even if I am a brain in a vat or the dream of a Martian, from a teleofunctionalist perspective phenomenal self-representata are only given in the historical context of a generating system." [Metzinger 2003: 278.]

There are two arguments in play here. The first is that the existence of a physical world is a fundamental assumption of naturalist theories of mind, the second the idea that if there is some construct (such as the VW) there also has to be something on the basis of which the construct is constructed.

The difficulty with the first argument is that it is not much of an argument at all, merely a statement of a key naturalist assumption. The second argument, on the other hand, does not assert that the world outside of the VW has to be in a certain way (that it is physical or divine) but simply claims that there has to be *something* other than the VW which brings the VW about. Yet this very claim is itself in need of support, since its truth is not obvious

Suppose someone argued that the world cannot just consist of sets. Since everything has to be a set of something generating it (its members) there has to be something inside every set. Even if we ignore the empty set this argument is still deficient, since it is only successful if we assume the truth of the axiom of foundation, but we know that there are perfectly functional versions of set theory that do not assume this axiom. (One way in which this argument for the existence of non-sets could fail if all chains of set-membership looped back on themselves.) What is required here is an argument why the "depends on" relation that has the VW as an antecedent has to be well-founded. In general the prospects for establishing such ontological foundation claims do not seem bright.

We can distinguish two different accounts of what happens at the right-hand side of the veil of perception available to the VW theorist. The picture supported by Metzinger's first argument falls under what we are going to call the strong account, that supported by the second falls under the weak account. The strong accounts believes there to be a considerable correspondence between the VW and the world outside, while the weak account does not want to postulate more than the bare *existence* of such a world.

The weak account does not reject the reality of the right-hand side, but denies that there is much we can say about it, apart from that it is there. There is no way in which we can access it directly, without the interface of the VW, and for this reason any claims about how successful the model is in "getting it right" are ill-founded. The existence of a "real world" is not under dispute, but this world is not something to which our familiar epistemic concepts of accurate and inaccurate representation could be applied.

It is apparent, however, that there are difficulties with both the strong and the weak account. The problem with the former is the presupposition that we could adopt a neutral perspective on the world that allowed us to access the accuracy of our representation in an objective way. But there are also reasons to doubt the cogency of postulating a largely undefined "I-know-not-what" on the right-hand side, as proposed by the weak account. This doubt finds its main support in the view that our all conceptual resources derive their function and purpose from their role in the VW. The word "all" here includes the existential quantifier and the notion of an existential dependence relation, and for this reason talk of an independent something everything in the VW depends on can only be used and understood within the VW and not, as the weak account would assume, by a suitable connection to something outside of the VW.

Note that this position (which we will refer to as irrealist) is not a nihilist view claiming that there is nothing on the right-hand side. For saying there is nothing encounters exactly the same problem as saying there is something, namely that of subsuming the right-hand side external to the VW under the conceptual framework that constitutes the VW. Rather, the irrealist would want to say that the right-hand side exists only as part of the VW.

The irrealist relies on three main arguments against the strong and weak interpretation of VW theory.

1. The consistency argument

The aim of the first argument is to refute the "variant of the Cartesian cogito" asserting that, given the existence of the VW, some constructing system has to exist. It cannot establish irrealism but sets out to show that the foundationalist viewpoint expressed in the strong and weak account is not the only possible interpretation of VW theory. It points out the difficulty the foundationalist positions have in excluding the possibility that the "depends on" relation that has the VW as a first relatum is circular. If we accept the idea of the noumenal as a mere ens rationis, having followed the dependence relation down up to whatever "extracranial" or "extradermal" reality has generated it we find ourselves once again in the confines of the VW, since this reality is itself part of the VW. If this argument succeeds, that is if the circular interpretation is consistent, it will have been shown that -- in the absence of other considerations -- the irrealist's interpretation is one possible interpretation of VW theory besides the foundationalist accounts.

2. The conceptual map argument

The second argument sets out to provide more direct support for the irrealist interpretation by arguing against the coherence of the standpointless view the foundationalist readings seem to presuppose. The key point is the observation that for the VW theory all conceptual resources are on a par since they are all part of the VW. It is therefore difficult to argue that certain resources (such as colour concepts) are only part of the VW, while others (such as "object" and "property") also apply beyond the VW. [This is the difficulty Strawson [1992: 64.] has in mind when he points out that a claim of a correspondence between perception and the world perceived cannot be cashed out as "an invitation to step outside the entire structure of the conceptual scheme which we actually have-and then to justify it from some extraneous point of vantage. But there is nowhere to step; there is no such extraneous point of vantage."]

The foundationalist supporting the strong or weak interpretation wants to argue that on the one hand there is our conceptual map, which allows us to successfully interact with the world, and on the other hand there is the real terrain, whatever the conceptual map is a map of. Yet the irrealist will point out that in VW theory map and terrain coincide: we use the concepts to move through a virtual world the very structure of which is constituted by the concepts. To assume, as the weak interpretation does, that there is something beyond the map, something which cannot be captured by our conceptual resources is like postulating the existence of a novel chess opening so sophisticated that it cannot be captured by the resources of chess notation.

## 3. The self-application argument

VW theory holds that all human cognition takes place in the brain-based representation that is the VW. This then raises the question what precisely the status of VW theory is by the standards of VW theory. For we seem to be facing the dilemma that either the thought that there is a world outside of the VW is not a human cognition, or that this thought only takes place in the VW.

The claim that it is not a *human* cognition could be cashed out by arguing that only a divine knower can know the truth that there is something behind the VW, but that our only way of epistemic access to this truth is by faith in the assertion of the divine knower. However, appeal to divine cognition is not really compatible with the naturalist framework espoused by VW, so this hardly seems to be the way the defender of the strong or weak interpretation of VW theory wants to go. In addition, arguing that the thought of the existence of a VW-independent world is not a cognition at all (human or divine) appears to be difficult to accept as well; the statement "there is something beyond the VW" certainly looks like a knowledge claim.

So we are left with the second option, namely that the thought that there is a world outside of the VW is part of the VW. But all parts of the VW that can be evaluated for truth get their truth-value relative to the VW. The statement that there is a red apple in front of me is true if the VW I am located in contains a red apple in a suitable spatial relation relative to me. Yet if the claim "there is a world outside of the VW" is evaluated in this way, and comes out as true, the only reason for this can be the irrealist one, namely that the "world outside" is part of the VW.

We are therefore left with a hypothetical conclusion: if one wants to adopt VW theory, the best account of the relation between mind and world is an irrealist one. Whether it is desirable to accept VW theory in the first place is of course a different question.

## References

Richard Dawkins 1999: Unweaving the rainbow: science, delusion, and the appetite for wonder, London, Penguin.

Thomas Metzinger 2003: Being No One : The Self-model Theory of Subjectivity, Cambridge, Mass. : MIT Press.

-- 2010 The Ego Tunnel: The Science of the Mind and the Myth of the Self, New York : Basic Books.

P.F. Strawson 1992: Analysis and Metaphysics : An Introduction to Philosophy, Oxford : Oxford University Press.